MINISTRY OF HEALPH OF REPUBLIC OF BELARUS
VITEBSK STATE MEDICAL UNIVERSITY

Козлов А.И.

БИОЛОГИЧЕСКАЯ ФИЗИКА

BIOLOGICAL PHYSICS

for the 1st – year foreign students of a pharmaceutical faculty

Рекомендовано учебно-методическим объединением по высшему медицинскому, фармацевтическому образованию Республики Беларусь в качестве пособия для студентов учреждений высшего образования, обучающихся по специальностям
1-79 01 08 «Фармация»

Витебск, 2021

УДК 57+53=111
ББК 28.071я73
К80

Рецензенты:

И.Ф.Кашевич, доцент кафедры инженерной физики Витебского Государственного Университета имени П.М.Машерова, к.ф.-м.н.;

Кафедра технической физики Белорусского Национального Технического Университета, заведующий кафедрой доцент И.А. Хорунжий, к.ф.-м.н.

Козлов, А. И.
К80 Biological physics = Биологическая физика : lectures for the 1st-year students of a pharmaceutical faculty : пособие / Козлов А. И. – Витебск: ВГМУ, 2021. – 191 с.

ISBN 978-985-580-070-6

The manual “Biological physics” is prepared in accordance with the educational program on physics for students of pharmaceutical faculties of medical universities.

УДК 57+53=111
ББК 28.071я73

ISBN 978-985-580-070-6

CONTENTS

INTRODUCTION

Lecture course “Biological Physics“ is included in the academic curriculum of the first-year students of pharmaceutical faculties of medical universities. Physics is an essential part of education of future pharmaceutists, because physical concepts and methods adopted from the theory of oscillations, statistical physics, thermodynamics and quantum mechanics form foundation of chemistry. In addition, basic knowledge in electricity, optics, radioactivity are also necessary for understanding of working principles of laboratory equipment.

The proposed manual «Biological physics» is written on basis of lectures delivered to the first-year students of Vitebsk State Medical University. This course is in agree with the state educational program for the speciality 1-79 01 08 “Фармация“ (“Pharmacy”). The main purpose of this manual is to tie students’ school knowledge of physics with demands of chemical chairs and provide future pharmaceutical chemists with necessary basis. The main difficulty in this way is wide difference of school programs on physics and mathematics in different countries and thus following essential diversity of initial students’ background. Therefore the proposed course is based only on elementary algebra (concept of a function, its graphical representation, algebraic equations of the first and second order, concept of a vector). It is assumed that necessary information on differential and integral calculus, differential equations of some kinds and on probability are should be taught in the parallel course of mathematics for pharmaceutical specialities. However some information on vector products and differential equation beyond the mathematical program is given in the course of the proposed manual.

The manual contains 14 lectures according to the state educational program. Some simple results following from the lectures are formulated as problems for self-education.

The author hopes that study of physics (and necessary mathematics) for pharmaceutical chemists will not come to the end after the final exam but will be continued later in case of need.

LECTURE 1. ROTATIONAL MOTION

1. Kinematics and dynamics of translational motion

In the beginning we consider motion of very small objects (you can simply regard them as points with mass) under the action of different forces. Everybody knows from the school course of physics that there are so-called **inertial** frames. Their main characteristic feature is that any mass point stays in rest or moves along a straight line in such frame if sum of all forces acting at this point is null. It was sir Isaac Newton who first understood it therefore the content of previous two phrases is usually called **the first Newton's law** of mechanics. The Earth moves around the Sun along approximately circular orbit and rotates around its own axis therefore strictly speaking the frame associated with our planet is not inertial one but later we will make sure that it could considered as inertial with some reasonable degree of accuracy.

Location of every moving object at any moment of time can be described with its displacement from the reference point as a function of time, for example $x(t)$. Then instant **speed** of the object at every moment is defined as the derivative of the displacement on time:

$$v(t) = \frac{dx}{dt} \tag{1}$$

Movement of an object takes place with a constant speed very rarely. Indeed, there should be some initial interval of time when the object gathered its speed and in some time it would come to rest. It means that to describe processes of speed gathering or slowing-down we need to introduce the speed of change of speed or concept of acceleration. **Acceleration** of an object is the first derivative of its velocity on time or the second derivative of its displacement:

$$a(t) = \frac{dv}{dt} = \frac{d^2x}{dt^2} \tag{2}$$

Of course, acceleration also can vary with time, therefore we denoted it in the latter expression as a function of time.

Newton had not stopped at formulation of his first law and found additionally (among many others his scientific achievements) that acceleration of a moving object in any inertial frame is proportional to sum of all forces acting at it and coefficient of proportionally is equal to inverse mass of the object:

$$a(t) = \frac{F}{m} \tag{3}$$

The latter formula offers the famous **second Newton's law** of motion. Letter m denotes the mass of an object. The more is its mass the less is its acceleration as it follows from Eq. (3), so we say that mass is a measure of **inertial** properties of the object, or ability of the object to resist to change of its speed.

Study of translational motion of any object is in fact based at formulation, solution and analysis of equation (1). It should be also emphasized that equations (1)–(3) are valid if the net force acting at the point is directed along *x*-axis and as a result translational motion of that point takes place along the same axis. As you know from the school course of physics, in general case the vector formulation of all equations mentioned above is necessary.

There is another one very important concept of mechanics called **momentum**. Momentum of a point object is the product of its mass and velocity:

$$p = m \cdot v \tag{4}$$

The second Newton's law of motion can be rewritten using momentum in the next way:

$$F = \frac{dp}{dt} \tag{3,a}$$

This is a more general formulation of the main law of dynamics, than (3). Formula (3, a) is often used in the theory of relativity as well as in quantum mechanics.

It is seen from Eq. (3, a) that if sum of forces acting at an object is equal to zero, than its momentum remains constant in time (recall properties of derivatives, please). The last sentence is a formulation of one of basic physical principles – **the law of conservation of momentum**. I hope you have heard about another one very important principle – **the conservation of energy law**. One of its simplest possible formulations is: if values of all forces acting at an object depend only on distances between this object and other objects surrounding it, then total energy of the object remains constant. The famous German mathematician Emma Noether (more exactly, Nöther in German) had proven a theorem in 1918, which stated that all conservation laws are corollaries of some symmetries in our Universe. So

she had stated that momentum conservation is the consequence of equivalence of all points in space (physical laws in all points are the same), energy conservation is the consequence of equivalence of all moments of time after the Big Bang (physical laws remain the same in time) and so on.

2. Operations of vector multiplication

As you know a **vector** is a piece of a straight line ended with an arrow, which indicates the direction of the vector. Any physical quantity, which is considered as a vector, is characterized not only with its relative value but with its direction too. So the point's displacement, velocity, acceleration as well as the force considered in the previous paragraph – all they are vector quantities.

Addition and subtraction of vectors are usually considered at school. Now we consider two more vector operations.

Dot product or **scalar product** of two vectors is a number, which is equal to the product of moduli of both vectors and of cosine of angle between them (*Fig. 1*, below). In fact, dot product is a projection of one of the vectors at the direction of another one:

$$p = \vec{a} \cdot \vec{b} = |\vec{a}| \cdot |\vec{b}| \cdot \cos\alpha \qquad (5)$$

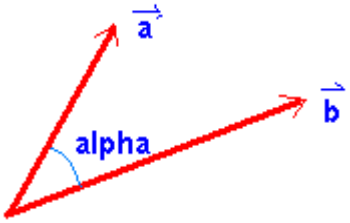


Fig. 1

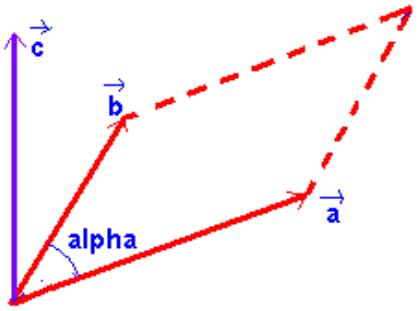


Fig. 2

Cross product or **vector product** of two vectors is a vector, which is equal to the product of moduli of both vectors and sine of angle between them (*Fig. 2*) and is directed from an observer if he/she is able to rotate $\vec{a}$ to $\vec{b}$ at smaller angle clockwise (in opposite case it is directed inversely):

$$\vec{c} = \vec{a} \times \vec{b} = |\vec{a}| \times |\vec{b}| \cdot \sin\alpha \qquad (6)$$

In fact, the absolute value of the cross product of two vectors is

equal to area of a parallelogram made up at those vectors as its sides. It is easily seen from expression (6) that the vector product of two vectors is zero if they are collinear (parallel to each other). Operation of cross product of two vectors is not commutative, that is:

$$\vec{c} = \vec{a} \times \vec{b} = -(\vec{b} \times \vec{a})$$

As the scalar product of two vectors is the operation between numbers (vectors' lengths) it is commutative.

3. Kinematics and dynamics of rotational motion

Now we proceed with our consideration with more complicates objects – solids (consisting of vast amount of small mass points). We start to study rotation of a body. Everyone knows that rotation is a kind of motion in which all points of a body circumscribe circles around a common axis. As in the case of the translational motion the path passed by a point in rotational motion can be described by its displacement as a function of time $x(t)$. But such approach demands description of motion of infinite number of mass points. However, it is much easier to measure an angle of rotation depending on time: $\varphi(t)$, because this angle is the same for all points of the body except those lying at the axis of rotation. Thus to characterize instant speed of rotation at any moment of time the derivative of the rotation angle on time is used, called **angular speed**:

$$\omega(t) = \frac{d\varphi}{dt} \tag{7}$$

Analogously speed of change of angular speed is called **angular acceleration** and it can be found as the second derivative of the angle of rotation:

$$\alpha(t) = \frac{d\omega}{dt} = \frac{d^2\varphi}{dt^2} \tag{8}$$

It is also possible to find instant translational (or linear) velocity and acceleration of any rotating point:

$$\vec{v}(t) = \vec{\omega}(t) \times \vec{r} \tag{9}$$

The expression for angular acceleration has the simplest form if axis of rotation is fixed:

$$\vec{a}(t) = \vec{\alpha}(t) \times \vec{r} = \vec{\omega} \times (\vec{\omega} \times \vec{r}) \qquad (10)$$

where $\vec{r}$ is the position vector or radius-vector of a rotating point (it is drawn from the origin of coordinates to the point). In formulas (9)–(10) angular speed and acceleration are denoted as vectors. Indeed, consideration of values of rotational angle, speed and acceleration can simplify many formulas of rotational motion (*Fig. 3*). The next agreement about directions acts in this way:

1. *vector of the angle of rotation is directed along the axis of rotation so that we could watch rotation counterclockwise when "sitting" at the vector's arrow;*
2. *direction of the vector of rotational speed coincides with the direction of the angle of rotation;*
3. *direction of the vector of rotational acceleration coincides with the direction of the vector of rotational speed if that speed grows and is opposite to the speed vector if it diminishes.*

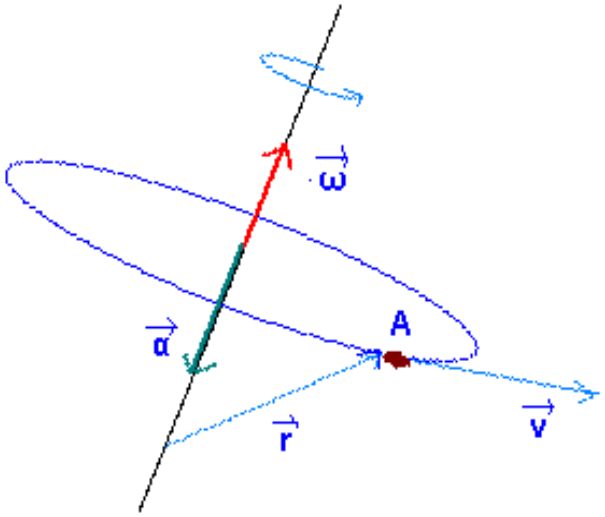


Fig. 3

Thus, at *Fig. 3* we can see a rotating point *A* and vectors of linear and angular velocity and of angular acceleration (as it is directed opposite to the angular speed, we can conclude that rotation becomes slower). Could you guess what is the direction of a vector of linear acceleration of the point *A*?

Now as we are familiar with concepts of angular velocity and acceleration and can describe rotational motion of a body at any moment of time we should consider the reasons of rotation. We know from our everyday experience that if any force acts at the axis of rotation so it moves translationally, it hardly influence angular speed of point *A*. You can recall a bicycle wheel: if it stays in rest and you try to move it holding its axis, it would move translationally, as a whole, all its point in this motion would pass the same distances. But if we push it at its rim in tangential direction, it would rotate. This example shows that the applied force should not pass through the axis to influence at rotation around it. Or more accurately we can say that a force passing through the axis of rotation does not affect revolution. Therefore a notion called **torque** is to be introduced here

$$\vec{M} = \vec{r} \times \vec{F} \qquad (11)$$

It is seen that absolute value of torque is proportional not only to the force applied to a mass point but also to its radius-vector (you can consider the same point *A* at the *Fig. 3*). Indeed, everybody knows the farther from an axis is a point at which one rotates a bicycle wheel the easier is to spin it. In terms of physical equations this finding can be written as the **law of dynamics of rotational motion**:

$$\vec{\alpha} = \frac{d^2\vec{\varphi}}{dt^2} = \frac{\vec{M}}{I} \qquad (12)$$

I hope you recognize on the left-hand part of Eq. (12) vector of angular acceleration and the torque on its right-hand side. It is easily seen that Equations (12) and (3) have the same mathematical form. It suggests that an unfamiliar so far value *I* in the denominator of Eq. (12) somehow plays the role of inertial mass.

4. Moment of inertia

Let us consider some body of irregular shape rotating around the axis *ZZ*’ with angular velocity ω (*Fig. 4*).

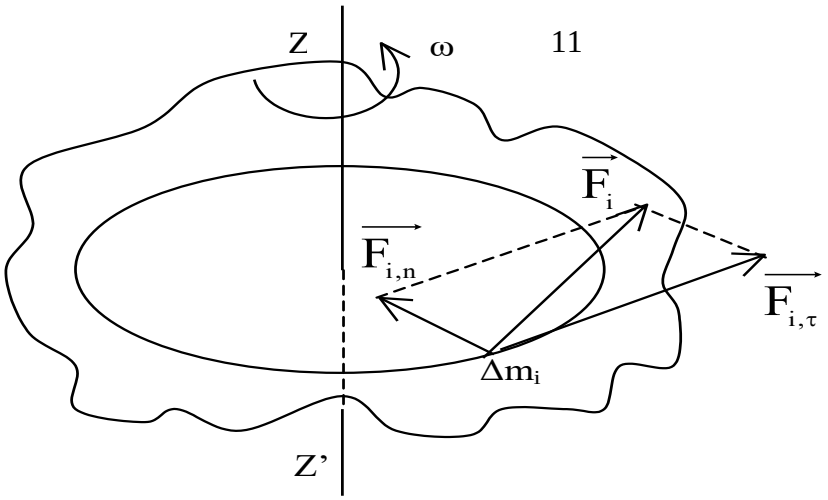


Fig. 4

Suppose a force $\vec{F}_i$ acts at some very small i^{th} element of that body (mass of the element is Δm_i). Any force is a vector quantity and every vector can be considered as a sum of its non-collinear components. So the force $\vec{F}_i$ can be represented as the sum of two components: $\vec{F}_{i,n}$ is the so-called normal component as it is directed from the element Δm_i to the axis of rotation, $\vec{F}_{i,\tau}$ is the tangential component. As parameters of rotation depend not on the force but on applied torque (according to Eq. (12)), we would consider only tangential component of external force $\vec{F}_{i,\tau}$. According to Eq. (3) and Eq. (10)

$$\vec{F}_{i\tau} = \Delta m_i \vec{a}_i = \Delta m_i \left(\vec{\alpha} \times \vec{r}_i\right) = \Delta m_i \left(\frac{d^2 \vec{\varphi}}{dt^2} \times \vec{r}_i \right)$$

as vectors $\vec{\alpha}$ and $\vec{r}_i$ are mutually perpendicular, we can write for their absolute values

$$\left|\vec{F}_{i\tau}\right| = \Delta m_i \cdot \alpha \cdot r_i = \Delta m_i \cdot r_i \cdot \frac{d^2\varphi}{dt^2}$$

therefore

$$F_{i\tau} \cdot r_i = \Delta m_i \cdot r_i^2 \cdot \frac{d^2\varphi}{dt^2} = \Delta m_i \cdot r_i^2 \cdot \alpha$$

Because the left-hand side of the latter equation is in fact a torque of the force applied to the i^{th} mass point of the body at *Fig. 4* then comparing this equation with Eq. (12) we can conclude that the **moment of inertia** of that element is equal to

$$I_i = \Delta m_i \cdot r_i^2 \tag{13}$$

The moment of inertia or the **second moment** of any object rotating around the some axis is equal to the sum of all infinitesimally small elements of that object calculated as in Equation (13):

$$I = \sum_i \Delta m_i \cdot r_i^2 = \int_{body} mr^2 dr \tag{14}$$

As the moment of inertia depends not only of mass of a body but also on remoteness of its different parts from an axis of rotation, it can be easily realized that the second moment I_{disc} of a flat disc with radius *R* and mass *m* is smaller than that of a ring of the same mass and radius:

$$I_{disc} = \frac{1}{2} MR^2 < I_{ring} = MR^2$$

Another one corollary of Equation (14): if some object can rotate around different axes, then its moments of inertia would be different. This is stated by the **theorem of Hyugens-Steiner**, namely

Moment of inertia of a body calculated with respect to any arbitrary axis is equal to sum of moments of inertia of that body calculated with respect to the axes passing through the center of mass plus the product of the mass of the body and the squared distance between two considered axes:

$$I_a = I_c + m \cdot r_a^2$$

For example, the second moment of a uniform thin rod of length *L* is $I_c = \frac{1}{12} m \cdot L^2$ if rotation takes place around its center of mass and for the case of rotation around its end is

$$I_a = \frac{1}{12} m \cdot L^2 + m \cdot \left(\frac{L}{2}\right)^2 = m \frac{L^2}{3}$$

5. Kinetic energy in the rotational motion

Suppose a solid rotates around some axis with angular speed ω (like at *Fig. 4*). A small element of this body with mass Δm_i has kinetic energy

$\Delta E_i = \Delta m_i \cdot v_i^2 / 2$, where v_i is the linear speed of motion of that element. Using (9) we can rewrite the latter expression in the next way:

$$\Delta E_i = \frac{\Delta m_i r_i^2 \omega^2}{2}$$

where r_i is the distance between the i^{th} element and the axis of rotation. Then total kinetic energy of all elements, of which the body is composed, is equal to the next sum

$$E = \sum_i \left(\Delta m_i \cdot r_i^2 \omega^2 / 2 \right) = \frac{\omega^2}{2} \sum_i \left(\Delta m_i \cdot r_i^2 \right) = \frac{I \cdot \omega^2}{2}$$

That is the kinetic energy of a rotating body is equal to half of product of moment of inertia and squared angular velocity.

6. Angular momentum

Let us consider a table of parallel values describing translational and rotational motion and their units of measurement in the SI-system:

Translational motion		**Rotational motion**	
Displacement	*x*, [*m*]	angle of rotation	φ , [*rad*]
speed	*v*, [*m/s*]	angular speed	ω, [*rad/s*]
acceleration	*a*, [m/s^2]	angular acceleration	α, [rad/s^2]
mass	*M*, [*kg*]	moment of inertia	*I*, [$kg*m^2$]
force	*F*, [*N*]	torque	*M*, [*N*m*]
momentum	*p*, [*kg*m/s*]	?	?

The last two cells of the latter table are vacant. Let us fill them, so we introduce a value called **angular momentum**:

$$\vec{L} = m \cdot \left(\vec{r} \times \vec{v} \right) \qquad (15)$$

In the case of a point mass object rotating around a fixed axis the later formula can be rewritten:

$$L = m \cdot r \cdot v = m \cdot \omega \cdot r \cdot r = I \cdot \omega \qquad (16)$$

The last result can be generalized for a body of arbitrary shape and size and even for a moving axis of rotation, (although we do not prove it in our course)

$$\vec{L} = I \cdot \vec{\omega} \tag{17}$$

A question: in what units is angular momentum measured? Fill the table above.

Consider derivative of angular momentum on time if the distance between a rotating object and an axis of rotation remains invariant ($\vec{r} = const$):

$$\frac{d\vec{L}}{dt} = \frac{d(m(\vec{r} \times \vec{v}))}{dt} = m\frac{d\vec{r}}{dt} \times \vec{v} + m\vec{r} \times \frac{d\vec{v}}{dt} = \vec{r} \times \frac{d\vec{p}}{dt}$$

If you recall now the second Newton's law of motion in form of Eq. (3, a), you would see that

$$\frac{d\vec{L}}{dt} = \vec{r} \times \vec{F} = \vec{M}$$

A simple but very important corollary follows from the latter equation: if overall torque of external forces acting at a rotating body is equal to zero then angular momentum remains constant according to properties of derivatives:

$$\frac{d\vec{L}}{dt} = 0 \Rightarrow \vec{L} = I \cdot \vec{\omega} = const$$

If fact, this is the **law of conservation of angular momentum**. It should be added also that this law is the consequence of equivalence of all directions in space.

7. Coriolis force

In the very beginning of this lecture we mentioned the first Newton's law of motion, which postulates existence of inertial frames. In fact, the second his law is valid only in such frames. Imagine a boy staying with his eyes closed at a rotating platform. He acts at the platform with the force of his weight and the platform resists with the equal force so he does

not move in vertical direction – quite in compliance with the second law. But in the direction parallel to the plane of the platform some force acts at him and causes him to change his location even if he does not move. It means that sum of all forces acting at him in the state of rest is not equal to zero. Why does the first law violate here? Because the boy's frame is not inertial. His platform rotates, or in other words moves with acceleration, which obeys Eq. (10). The second law has the next form in this case:

$$m \cdot \vec{a}_{in} = \vec{F} - m \cdot \vec{a}$$

where $\vec{a}_{in}$ is his acceleration with respect to the rotated platform, $\vec{a}$ is the linear acceleration in rotation, $\vec{F}$ is a possible external force independent of rotation. If we can neglect friction $\vec{F} = 0$ (as at the platform covered with ice) then boy's acceleration is equal and directed against the vector $\vec{a}$. In such case we usually say that some **inertial force** acts at the boy:

$$\vec{F}_{in} = m \cdot \vec{a}_{in} = -m \cdot \vec{a}$$

You could ask how Newton guessed about his second law of mechanics if he had never left our rotating Earth? Let us estimate acceleration of an object placed at any equatorial point due to Earth's rotation around its own axis. We use Equation (10) for the case of mutually perpendicular radius-vector and a vector of angular speed. Our planet's average radius is about 14000 *km*, so any point at the equator has inertial acceleration equal to

$$a = \omega^2 \cdot r \approx \left(\frac{2\pi}{24 \cdot 3600} \right)^2 \cdot 7000000 \approx 0{,}037 \frac{m}{s^2}$$

This quantity is about 300 times less than the gravitation acceleration. Inertial acceleration in the Earth's rotation around the Sun is even smaller (try to estimate its value). Therefore we usually neglect such inertial acceleration in our life.

Let us return to the boy rotating at the platform. If he tries to go from the center of the platform to its edge during rotation he could fill some other force besides inertial one. Inertial force acts in the direction to the edge independently on his motion along the platform. But the additional force, which becomes apparent if boy tries to go along the radius of the platform, is called the **Coriolis force**. It is equal to

$$\vec{F}_{Cor} = -2 \cdot m \cdot (\vec{\omega} \times \vec{v})$$

where $\vec{v}$ is the velocity of the boy's motion with respect to the platform. Coriolis force always curves the real path of moving at rotated frame object (*Fig. 5*). The boy trying to come up to the point *A* really reaches a point *B*.

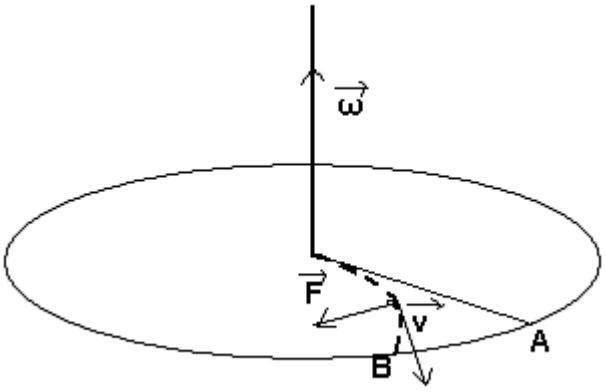


Fig. 5

7. Centrifugation

Concept of rotational motion is important in many branches of physics: in thermodynamics we usually consider different possible motions (including rotation) of molecules (degrees of freedom), some electrical and optical properties of matter are caused by rotation dipole molecules in the alternating electric field, in semiclassical theory of magnetism permeability of a substance depends on rotation of electrons, but there is also a direct application of rotation – separation of non-uniform substances like colloidal solutions, called **centrifugation**. There are many different kinds of centrifuges, they are used in chemistry, physics, engineering, agriculture, mining. Their common part is a rotating vessel for a liquid or gas mixture. In the absence of external forces at every part of a solution acts only inertial force, equal to

$$\vec{F}_{in} = m \cdot \vec{a}_{in} = -m \cdot \vec{a}$$

As the inertial force is proportional to weights of particles, heaviest ones have greater acceleration and therefore tend to move farther from the axis of rotation. As a result different components are separated at different distances from the axis depending on their mass. Usually the heaviest fraction is collected near the walls of the rotating vessel.

LECTURE 2. OSCILLATORY MOTION

1. Harmonic oscillations

Oscillatory motion or **oscillation** is such kind of motion of an object, which is characterized with repetition of its position with respect to some coordinate frame. Of course, rotational motion considered in the previous lecture obeys this definition but intuitively we can easily distinguish rotation and oscillation: speed of a rotating object does not change its direction in the object's frame while for oscillating object it does. The simplest kind of oscillation is called **harmonic** one, in this case displacement dependence on time obeys *sin* or *cos* function.

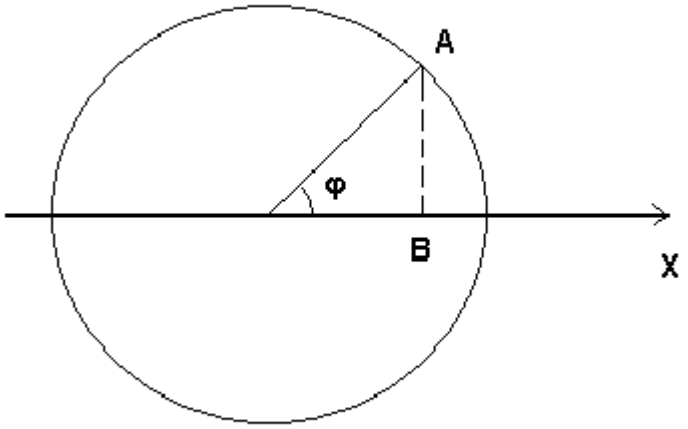


Fig. 1

For example, let us consider a point *A* rotating counterclockwise along circular trajectory, like it shown at *Fig. 1*. Its projection *B* at the horizontal axis *X* oscillates between the right and the left points of intersection of the circle with the axis *X*, instantaneous value of angle of rotation is denoted as φ. If rotation takes place with a constant angular speed ω, then time dependence of a horizontal coordinate of the point *B* is the next:

$$x(t) = x_0 \cdot \cos(\varphi) = x_0 \cdot \cos(\omega_0 t + \varphi_0) \qquad (1)$$

here the maximal value of displacement x_0 is called the **amplitude** of oscillation, φ_0 is its **initial phase**. According to the previously stated definition Equation (1) describes harmonic oscillation of the point *B*.

2. Differential equation of oscillations

Let us consider a mass *m* which can slide along a horizontal plane and which is connected to an elastic string. If the string is in equilibrium, that is it is neither stretched nor compressed, the mass is fixed at coordinate x_{eq} along the axis *X*, which coincides with the horizontal plane (*Fig. 2*). For simplicity we can consider $x_{eq} = 0$.

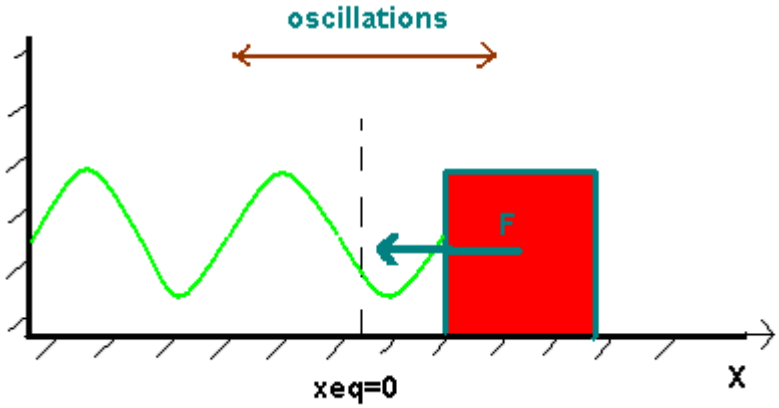


Fig. 2

If the string (shown with green wavy line) is stretched as it is drawn at the *Fig. 2* then an elastic force acts at the mass (the red rectangular) in the direction opposite to elongation of the string. Value of this force, according to the Hook's law, is proportional to the elongation $x = x - x_{eq}$:

$$F_{elast} = -k \cdot x$$

If the mass moves with some velocity $v = dx/dt$ then in general case a force of friction (of resistance) acts at it, proportional to the speed value:

$$F_r = -2 \cdot \gamma \cdot \frac{dx}{dt} = -2 \cdot \gamma \cdot v$$

Thus, from the second Newton's law of motion we have got

$$ma = -F_{elast} - F_r = -kx - 2\gamma v$$

As velocity is the first derivative of a displacement on time and acceleration is the second one, dividing the latter equation by the mass, we can easily get:

$$\frac{d^2x}{dt^2} + 2\beta\frac{dx}{dt} + \omega_0^2 x = 0 \qquad (2)$$

Expression (2) is the **equation of free mechanical oscillations**. There $\beta = \gamma/m$ is the **attenuation coefficient,** $\omega_0 = \sqrt{k/m}$ is called the **cyclic frequency of inherent oscillation.**

You have studied this kind of equations in the course of mathematics, so you know that (2) is called the linear ordinary differential equation of the second-order. To solve it, we have to compose formally an algebraic equation with the same coefficients:

$$y^2 + 2\beta y + \omega_0^2 = 0$$

then to solve the latter equation using the well-known Viète formula:

$$y_{1;2} = -\beta \pm \sqrt{\beta^2 - \omega_0^2}$$

The most interesting case for practical applications takes place if the expression under the radical is negative. In this case the solution of Eq. (2) looks like

$$x = \left(C_1 \cos\omega t + C_2 \sin\omega t\right) \cdot \exp\left(-\beta t\right) \qquad (3)$$

where C_1 and C_2 are constants of integration and $\omega = \sqrt{\omega_0^2 - \beta^2}$ is the **cyclic frequency of a damped oscillation**. It is easily seen that $\omega < \omega_0$. Now it becomes evident that the cyclic frequency ω_0 corresponds to absence of attenuation ($\beta = 0$) that is why it is called the inherent frequency – it concerns oscillation without any friction. Additionally, the value called **linear frequency** f is often used, which is 2π times less than an appropriate cyclic frequency. Linear frequency is inversely proportional to the period of oscillations T (that is the time interval in which one oscillation occurs):

$$f = \frac{\omega}{2\pi} \quad \text{and} \quad T = \frac{1}{f} \qquad (4)$$

Cyclic frequency is usually measured in *radian per second*, linear one – in *Hz*, while attenuation – in inverse *second* (although, in fact all this units mean inverse second).

Solution in form of Eq. (3) after some algebra can be rewritten as (check it yourself):

$$x(t) = x_0 \cdot \exp(-\beta t) \cdot \cos(\omega t + \varphi_0) \qquad (5)$$

Comparing equations (5) and (1), we can see that amplitude of damped oscillations decays in time according exponential dependence. To characterize rate of this decay another one value is used: **logarithmic decrement of attenuation**:

$$\lambda = \ln \frac{x(t)}{x(t+T)} = \beta T$$

3. Forced mechanical oscillations

Everybody knows from our everyday experience that any vibrational process ends sooner or later due to ubiquitous attenuation. Therefore to maintain vibrations we need to apply some external periodic force (recall seesaw). Then, the differential equation of oscillation has to include the constraining force on the right-hand side:

$$ma = -F_{elast} - F_r + F_c = -kx - 2\gamma v + F_0 \cdot \cos \omega_c t$$

or after division over *m*

$$\frac{d^2x}{dt^2} + 2\beta \frac{dx}{dt} + \omega_0^2 x = \frac{F_0}{m} \cos \omega_c t \qquad (6)$$

How does its solution look? In the simplest case of a periodic forced oscillation we can suppose that frequency of oscillations would be equal to the frequency of the applied force as after some time the steady state of oscillations would be established. Thus let us look the next expression for the steady solution in form of an oscillation with the frequency of the external force:

$$x = A\cos(\omega_c t - \varphi_c) \qquad (7)$$

where A is the amplitude of the constraint oscillation while φ_c is the phase shift between the applied force and the resultant oscillation (7).

Substitute Eq. (7) into Eq. (6) providing that

$$v = \frac{dx}{dt} = -A\omega_c \sin(\omega_c t - \varphi_c) \quad \text{and} \quad a = \frac{dv}{dx} = -A\omega_c^2 \cos(\omega_c t - \varphi_c)$$

so we have got

$$-A\omega_c^2\cos(\omega_c t-\varphi_c)-2A\beta\omega_c\sin(\omega_c t-\varphi_c)+A\omega_0^2\cos(\omega_c t-\varphi_c)=\frac{F_0}{m}\cos\omega_c t$$

This equation should be satisfied at any moment of time t in the steady state, therefore coefficients before $\cos(\omega_c t)$ and $\sin(\omega_c t)$ functions are independently equal to zero. We can find both unknowns A and φ_c in this way:

$$A=\frac{F_0}{m}\cdot\frac{\omega_0^2-\omega_c^2}{\left[\left(\omega_0^2-\omega_c^2\right)^2+4\beta^2\omega_c^2\right]\cdot\cos\varphi_c};\quad \operatorname{tg}\varphi_c=\frac{2\beta\omega_c}{\omega_0^2-\omega_c^2}$$

Eliminating trigonometric functions we obtain the expression for the amplitude

$$A=\frac{F_0}{m}\cdot\frac{1}{\sqrt{\left(\omega_0^2-\omega_c^2\right)^2+4\beta^2\omega_c^2}} \qquad (8)$$

It can be proven that A has maximum if $\omega_c=\sqrt{\omega_0{}^2-2\beta^2}$. This frequency is called **resonant** one and the phenomenon is called **resonance**. Resonance has numerous practical applications: tuning of radio- or TV-sets at a frequency of an appropriate station, lasers' operation as well as possible destruction of a bridge if the repetition rate of an external force (caused by a strong wind or synchronous periodic steps of military troops) randomly turns out to be equal to its inherent frequency.

4. Addition of oscillations of the same frequency

Let us solve the next problem first. Suppose, there are two oscillations which take place simultaneously in the same direction, they have identical frequencies and constant difference of initial phases. We suppose also that attenuation can be neglected for simplicity:

$$x_1=A_1\cos(\omega_0 t+\varphi_1) \qquad \text{and} \qquad x_2=A_2\cos(\omega_0 t+\varphi_2)$$

Adding these two oscillations and using some algebra for *cosine* of sum of two angles, we have

$$x = x_1 + x_2 = (A_1 \cos\varphi_1 + A_2 \cos\varphi_2)\cos\omega_0 t - (A_1 \sin\varphi_1 + A_2 \sin\varphi_2)\sin\omega_0 t$$

Let us introduce two new variables A and φ and suppose that

$$A_1 \cos\varphi_1 + A_2 \cos\varphi_2 = A\cos\varphi \quad \text{and} \quad A_1 \sin\varphi_1 + A_2 \sin\varphi_2 = A\sin\varphi$$

If to summarize squares of the latter equations and at the next step to divide the second of them over the first one (try to do it yourself as an exercise), both unknown variables would be found:

$$A = \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos(\varphi_1 - \varphi_2)} \qquad (9)$$

$$\operatorname{tg}\varphi = \frac{A_1 \sin\varphi_1 + A_2 \sin\varphi_2}{A_1 \cos\varphi_1 + A_2 \cos\varphi_2} \qquad (10)$$

Depending on initial phase difference $\varphi_1 - \varphi_2$ amplitude (9) of a resulting wave can have maximal (when $\cos(\varphi_1 - \varphi_2) = 1$), or minimal ($\cos(\varphi_1 - \varphi_2) = -1$), or some intermediate value ($-1 < \cos(\varphi_1 - \varphi_2) < 1$).

This example seems to be rather abstract but we will return back to formula (9) later in our course, because in fact we have considered now a very important phenomenon of **interference** of two waves.

5. Composite oscillations. The Fourier theorem

Small (in amplitude) oscillations of a mass point connected to a string or of a suspended pendulum can be harmonic, but in many cases like oscillations of a pendulum pulled off initially at angle exceeding 10^0 oscillations are evidently periodic but obey more complicated function than a harmonic one (*sin* or *cos*). Let us consider the next example. At the left half of *Fig. 3* time dependence of three periodic functions is drawn. All they have the same period $T = 2\pi$, the same amplitude ($x_0 = 1$), but their forms are obviously different. The upper graph corresponds to a sinusoidal (or harmonic, or monochromatic) wave, the middle is the rectangular wave, while the third one is the triangular wave. We can easily determine time dependence of velocity or any other parameter of the sinusoidal wave but how to describe analytically (by means of formulas) two other functions?

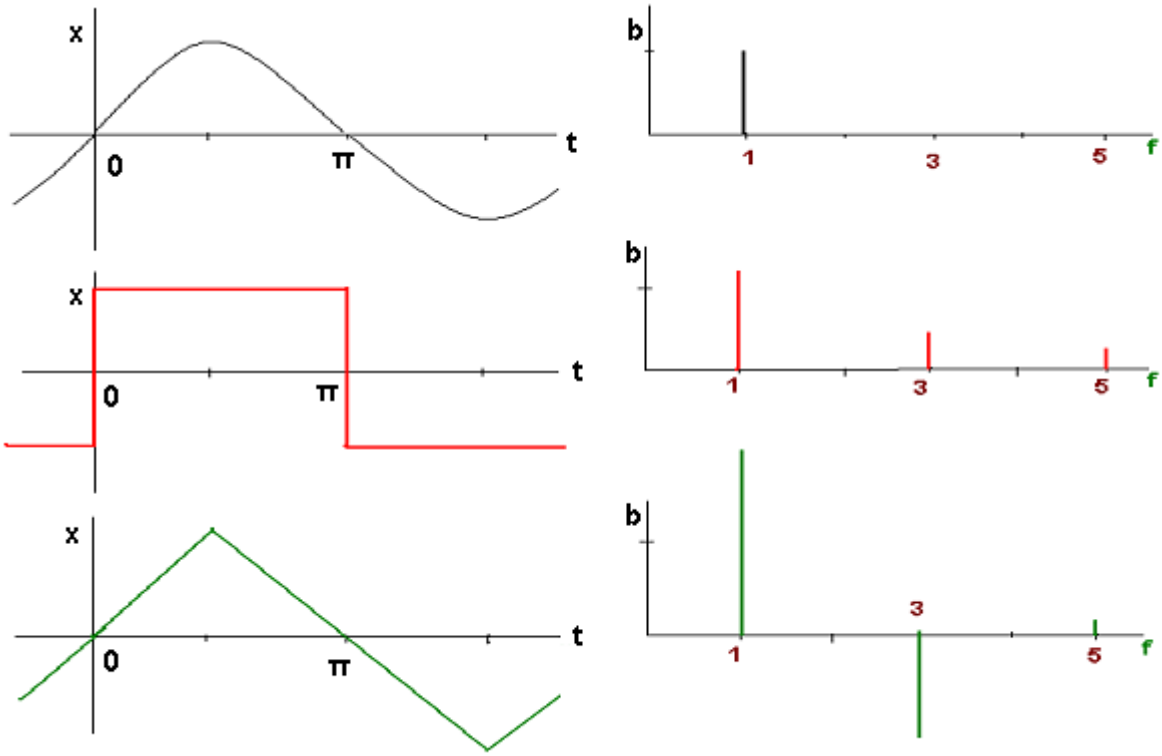


Fig. 3

The famous **theorem** named after **Fourier,** who had proved it for the first time, states:

Any periodic continuous function with period 2π *can be represented as a finite or infinite sum of sin and cos function in the next way:*

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cdot \cos(n\omega t) + b_n \cdot \sin(n\omega t)\right),$$

where coefficients of this series are equal to

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)dx \qquad (11)$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cdot \cos(nx)dx \qquad (12)$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cdot \sin(nx)dx \qquad (13)$$

It means that the rectangular wave from *Fig. 3* can be considered as

infinite series of waves with amplitudes obeying the formula:

$$f(x) = \frac{4}{\pi}\sin(\omega t) + \frac{4}{3\pi}\sin(3\omega t) + \frac{4}{5\pi}\sin(5\omega t) + \ldots,$$

Similarly, the triangular function's coefficients obey the next infinite series:

$$f(x) = \frac{8}{\pi^2}\sin(\omega t) - \frac{8}{(3\pi)^2}\sin(3\omega t) + \frac{8}{(5\pi)^2}\sin(5\omega t) - \ldots,$$

Appropriate coefficients can be found using formulas (11) – (13) for any function with period equal to 2π. In the case of arbitrary period change of variable *t* should be used.

6. Wave equation. Ultrasound.

Everybody knows that oscillations generated at some point in a continuous matter tend to spread around if properties of surrounding medium allow it. So a pebble thrown in a river excites **waves** transmitting vibration of water surface until a stone or sands at the riverside are met. Concept of waves is very important in the modern physics: acoustics and optics are devoted to study of propagated sound or electromagnetic waves. Quantum mechanics consider wave properties as indispensable for such objects as electrons, nucleons, atoms and so on.

Mathematical expression, which describes propagation of small-amplitude waves had been formulated in XVIII century. If u is the amplitude of displacement in a wave, its dependence on time and spatial coordinate is given by the **wave equation:**

$$\frac{\partial^2 u}{\partial t^2} = c_0^2 \frac{\partial^2 u}{\partial x^2} \qquad (14)$$

where u is called a displacement in general case (or pressure in the case of acoustic wave, or strength of electromagnetic field for optical wave), c_0 is the speed of propagation. Equation (14) is a bit different from the equation of oscillations (2) because in the former one the amplitude depends on two parameters – time t and distance x. Wave equation (15) also allows harmonic solutions like

$$u = u_0 \cdot \cos(\omega t - kx) \tag{15}$$

where u_0 is the amplitude of a wave propagating without attenuation in the positive direction of $x-$axis (from left to right), ω is the cyclic frequency, k is the wavenumber, which depends on the wavelength λ as follows: $k = 2\pi/\lambda$. Why can we assert that the wave obeying (15) propagates from left to right? Because form of sinusoidal profile does not change if the expression in brackets remains constant, that is only possible when *x* increases with increase *t*. It follows from Equation (14) that velocity of propagation of that wave is equal to

$$c_0 = \frac{\omega}{k} = \lambda \cdot f \tag{16}$$

It should be noted here that the amplitude of wave (15) cannot be determined from the wave equation. Knowledge of so-called **initial** or **boundary conditions** is necessary in this case.

You can easily check that expression

$$u = u_0 \cdot \cos(\omega t + kx) \tag{17}$$

is a solution of Equation (14) too. Harmonic waves (15) or (17) play very important role in science and applications, especially because Fourier theorem allows consideration of many different signals as results of summation of harmonic waves. Besides that in XVIII century a French scientist d'Alembert had found more general solution of the wave equation. He had stated that any continuous functions *f* and *g* satisfy equation (14) if their derivatives are also continuous:

$$u = f(x - c_0 t) + g(x + c_0 t) \tag{18}$$

I hope that you have guess that the function *f* corresponds to a disturbance propagating in positive direction of $x-$axis while *g* propagates oppositely (from right to left). You can easily check solution (18) substituting the second derivatives of amplitude on time and on the space coordinate in Equation (14).

Later we will consider propagation of electromagnetic waves, but it is necessary to mention here another one kind of waves surrounding us everywhere – sound or acoustic waves. Sound wave is the moving sequence of compressed and expanded parts in an elastic medium. Local pressure and density in those fields are slightly different compared with equilibrium state of the medium, for example, harmful value of acoustic pressure, which

damages human's eardrum, is about 60 *Pa*. Compared with normal atmospheric pressure equal to more than 10^5 *Pa*, this is an insignificant value. Acoustic waves cannot exist in vacuum, because there are no elastic forces between neighboring parts of space in the absence of a matter. As human ear can detect acoustic waves in rather narrow interval of frequencies from 16-20 *Hz* to 17-20 *kHz*, all acoustic waves are usually referred to one of three ranges: infrasound (frequencies less than 16 *Hz*), audible sound (above mentioned frequency range) and ultrasound (frequencies exceeding 20 *kHz*). Ultrasound has many medical applications in visualization of inner tissues. Low-intensity ultrasonic wave is not harmful for human health as compared with *X*-ray investigations. Besides that it was found that application of ultrasound to chemical reagents sometimes could increase rate of chemical reactions. This is the subject matter of **sono-chemistry**.

Advantage of these applications of ultrasound in comparison with infrasound or audible sound is in its higher frequency. It means that it has smaller wavelength according to Eq. (16). As the angle of diffraction divergence of a wave is of the order of (diffraction phenomenon will be considered later in optics):

$$\alpha \sim \frac{\lambda}{D} \tag{19}$$

where *D* is a character size of an obstacle, which is round by the wave (as at *Fig. 4* below). It follows from the latter formula that the less is the wavelength λ, the less important is diffraction effect, the sharper is geometric shadow of the obstacle. (In medical applications an "obstacle" means a bone, or a tissue, or external inclusion.)

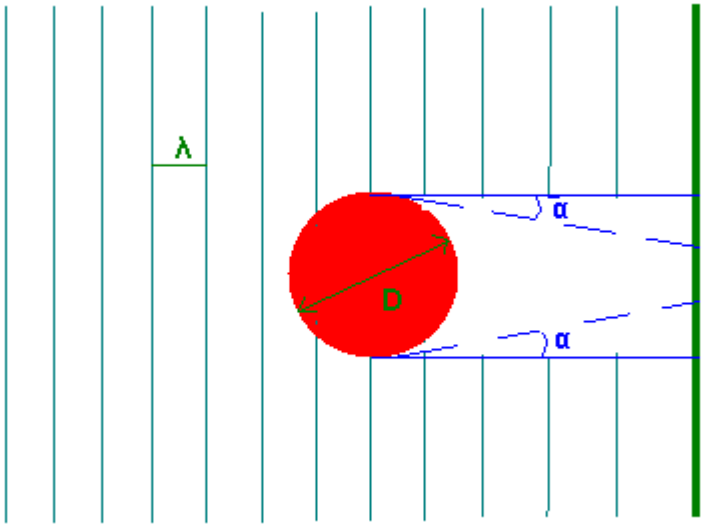


Fig. 4

7. Energy of waves

As acoustic wave propagates through a medium, oscillatory motion spreads in it. Any small volume of the medium, which takes part in oscillations, has some kinetic energy, proportional to square of its speed. Small parts of elastic medium, which left their position of equilibrium, have also some potential energy. It leads to conclusion that propagation of an acoustic wave means propagation of energy. All waves transmit energy – not only sound waves.

Quantitatively it is described with a concept of **energy flux**, which is equal to energy transmitted with a wave in a unit time:

$$\Phi = \frac{dE}{dt}$$

Energy flux is measured in *Watt*. **Intensity** of a wave (or density of power flow in it) is a power flow through a unit surface area perpendicular to direction of propagation of a wave:

$$I = \frac{dW}{dS}$$

Intensity of any wave is measured in *Watt* per m^2. As any wave carries energy it provides some pressure. This effect is rather weak although it was experimentally measured. Light-wave pressure had been first measured by Peter Lebedev in 1912 using very delicate experiments. Nowadays, engines using light pressure for interplanetary travel are often mentioned in science fiction.

8. Doppler effect

Let us imagine a whale, which is lying on the calm surface of ocean and strictly periodically splashing through the water with frequency *f*. As a result a wave propagates from the whale's tale along the surface of the ocean (*Fig. 5, a*) with the speed *c*. This wave is a periodic sequence of crests and troughs and the distance between neighboring crests is its wavelength λ.

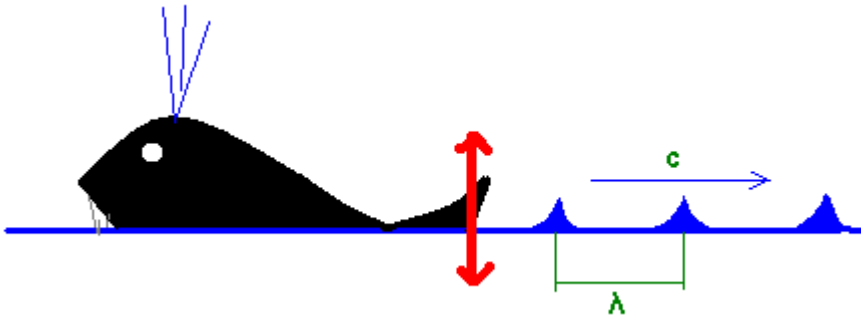


Fig. 5, a

Doppler effect (named after Christian Doppler who first discovered this phenomenon theoretically in the middle of XIX century) is a consequence of the fact that velocity of propagation of any wave in any substance depends on the wave's kind and of the substance's properties. Therefore if that whale slowly moves away from the initial point and continues the splashing with the same frequency and the same strength (*Fig. 5, b*) the wave departs with the same speed c. But the wavelength λ' is changed, because the whale moves and distance between crests evidently becomes longer.

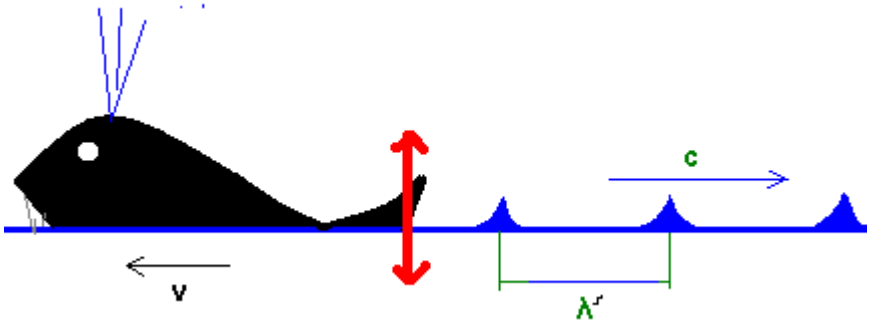


Fig. 5, b

If period of splashing is $T = 1/f = \lambda/c$ and the speed of whale's motion is v then

$$\lambda' = \lambda + v \cdot \tau = \lambda + \lambda \cdot \frac{v}{c}$$

So increment of wavelength because of the whale's motion with respect to observer is equal to

$$\Delta\lambda \equiv \lambda' - \lambda = \lambda \cdot \frac{v}{c} \qquad (20)$$

The wavelength increases, but frequency decreases from observer's point of view in this case. Thus it could be easily checked that

$$\Delta f \equiv f' - f = -\frac{\frac{v}{c}}{1+\frac{v}{c}} \qquad (21)$$

It is clear that if the whale switch "reverse gear" and move in opposite direction continuing its splashing then increments of the wavelength and of frequency change their signs.

Doppler effect becomes apparent also if the source of waves stays in rest while receiver moves. This phenomenon is widely used for measurement of speed of moving objects.

9. Nonlinear waves

All results, which are described in previous 8 sections of this lecture, are devoted to **linear** oscillations and waves. Equations (2) and (14) are also linear and their solutions obey **linear properties**. It means that if we multiple any solution of such equation by any non-zero constant, the resultant function would be also a solution of the same equation. Addition of two or more solutions of a linear equation is also the solution of the same equation. But many experiments showed that linear theory of waves satisfactorily describes real processes only if amplitudes of considered oscillations or waves are very small, in fact infinitesimally small. In a real life we often tend to make light brighter, sound louder and as a result we have to know how to consider oscillations with appreciable amplitudes. It turned out that equations (2) and (14) are not capable to describe such phenomena well enough. Besides that, since the second part of XIX physicists and engineers have discovered experimentally some phenomena (solitons, kinks, shocks), which could not obey linear equations of oscillations and waves. Therefore some brief information about such objects seems to be necessary for future pharmaceutical chemists. Propagation of **solitons** (or solitary waves) was first described by the famous Korteveg-de Vries equation:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \frac{\partial^3 u}{\partial x^3} = 0 \qquad (20)$$

where u is a displacement of particles of a substance in a soliton, t and x

are time and the spatial coordinate respectively. This is an essentially nonlinear equation because its second term contains the product of the unknown function u and of its derivative. Thus solutions of equation (20) do not possess the above mentioned linear properties. The soliton as the best-known its solution obeys the next expression:

$$u = \frac{3v}{\sinh^2\left[\frac{\sqrt{v}}{2}(x - vt)\right]} \qquad (21)$$

where v is the speed of propagation, *sinh* is the **hyperbolic sinus** – a function defined in the next way:

$$\sinh(x) = \frac{e^x - e^{-x}}{2}$$

It is seen from expression (21) that the more is the amplitude of a soliton, the more is its speed v and vice versa. This property was observed experimentally: a higher soliton overtakes the lower one and passes through it, whereupon continuing its motion.

In XX century different investigations had revealed that many processes, which had been previously considered as waves (nerve pulses, for example), turned out to be solitons.

LECTURE 3. MOLECULAR PHYSICS OF GASES

1. Main assumptions of the theory

The scope of studies of molecular physics embraces phenomena, which could be explained considering motion and interaction of molecules. Classical molecular physics is based on the next fundamentals:

1. *all substances consist of microscopic particles (atoms or molecules) and those particles move continuously and chaotically;*
2. *every small volume of a substance considered in molecular-kinetic theory contains very large amount of particles;*
3. *average distances between particles are much more than their sizes;*
4. *collisions between particles are considered to be perfectly elastic, without loss of kinetic energy;*
5. *particles are spread on all volume uniformly if external forces are absent;*
6. *modulus of velocity of any moving particle can have values from zero to infinity.*

First we consider a problem of determination of pressure of a gas in a vessel.

2. Basic equation of the molecular-kinetic theory of gases

As the total volume of gas molecules in a vessel is negligibly small compared with volume of the vessel (according to the 3rd preposition above) then every molecule moves freely major part of time suffering rare elastic collisions with other molecules. When colliding with a wall every molecule acts on the wall with some moment of force $\Delta F = f\Delta t$. Every small part of surface of the wall is bombarded uninterruptedly with a huge amount of molecules, this continuous chaotic action by molecules on the wall could be detected as gas pressure. It acts normally to the wall's surface at any point.

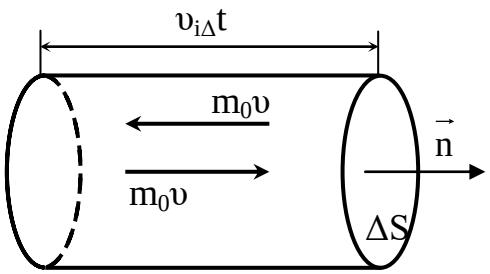


Fig. 1

Consider an arbitrary small round element of area of the wall ΔS (*Fig. 1*) and let us count number of molecules collisions with the element in time Δt. Evidently, it is equal to the number of molecules, which could reach the considered element in time Δt. Molecules can have different values of velocities acting in different directions. Besides that those

velocities chaotically and uninterruptedly change in collisions. However, if external conditions remain constant, total sum of kinetic energies of molecules remains constant too. Energy is redistributed but not absorbed in elastic collisions. Because of huge number of processes of this kind we could consider distribution of values of molecules' speeds to be permanent.

Any motion in 3*D* space could be described in a rectangular reference frame consisting of three mutually perpendicular axes *OX, OY, OZ*. Any direction of motion is equally probable for any molecule. So it could be assumed that 1/3 part of molecules moves along any coordinate axis and half of this amount (1/6 of total number) moves in the positive direction of the corresponding axis while another 1/6 part moves in the opposite (negative) direction.

Let us apply this reasoning to a circular straight cylinder with base ΔS (*Fig. 1*). Number of molecules ΔN_i with speed v_i **,** which collide area element ΔS in time Δt is

$$\Delta N_i = \frac{1}{6} n_i \Delta S v_i \Delta t \qquad (1)$$

where n_i is the concentration of molecules of the kind *i*, $\Delta S \cdot v_i \cdot \Delta t$ is the volume of a cylinder containing molecules of the kind *i*, which are capable to reach element ΔS at time Δt**.** When colliding with the wall, each molecule of mass m changes its momentum on magnitude, because in elastic collision velocity of a molecule retains the same value, but change its sign (direction of propagation):

$$\left|\Delta(mv_i)\right| = \left|-mv_i - mv_i\right| = 2mv_i \qquad (2)$$

Then according to the third Newton's law of motion, impulse gained by the wall in every collision is equal to $F_i \Delta t = 2mv_i$, but the impulse transmitted by molecules of the kind *i* to the area element ΔS in time Δt is

$$F_i \Delta t = 2mv_i \Delta N_i = \frac{1}{6} n_i \Delta S \Delta t v_i \cdot 2mv_i = \frac{1}{3} n_i m v_i^2 \Delta S \Delta t = \frac{2}{3} n_i \frac{m v_i^2}{2} \Delta S \Delta t$$

or

$$F_i \Delta t = \frac{2}{3} n_i \varepsilon_i \Delta S \Delta t \qquad (3)$$

here $\varepsilon_i = \frac{m\upsilon_i^2}{2}$ is the kinetic energy of translational motion of a single molecule.

The total impulse transmitted to the surface element ΔS in time Δt by **all** kinds of molecules in the vessel obeys the next expression

$$F\Delta t = \frac{2}{3}n_1\varepsilon_1\Delta S\Delta t + \frac{2}{3}n_2\varepsilon_2\Delta S\Delta t + ... = \sum_{i=1}^{\hat{e}}\frac{2}{3}n_i\varepsilon_i\Delta S\Delta t = \frac{2}{3}\Delta S\Delta t\sum_{i=1}^{\hat{e}}n_i\varepsilon_i \quad (4)$$

Dividing both parts of Equation (5) by $\Delta S \cdot \Delta t$, we obtain the next formula for gas pressure p

$$p = \frac{F\Delta t}{\Delta S\Delta t} = \frac{F}{\Delta S} = \frac{2}{3}\sum_{i=1}^{\hat{e}}n_i\varepsilon_i \quad (5)$$

where $\sum_{i=1}^{\hat{e}}n_i\varepsilon_i$ is the total kinetic energy of molecules placed in a unit volume or bulk density of the kinetic energy of the molecules' translational motion. Let us denote $\sum_{i=1}^{\hat{e}}n_i\varepsilon_i = w$**.** Then (5) looks like

$$p = \frac{2}{3}\ w \quad (6)$$

Expression (6) is **the basic equation of the molecular-kinetic theory of a perfect gas (Clausius, 1857)**. It states that

pressure of a perfect gas is equal to 2/3 of kinetic energy of translational motion of molecules occupying a unit volume.

Another form of equation (7) could be obtained if one uses a value of average kinetic energy of a molecule of a perfect gas as $\bar{\varepsilon} = \sum_{i=1}^{N}\varepsilon_i \Big/ N$**.** Then $w = n\bar{\varepsilon}$ and (6) takes on the next form

$$p = \frac{2}{3}n\bar{\varepsilon} \quad (7)$$

But the mean kinetic energy of a molecule is expressed through its speed

$$\bar{\varepsilon} = \frac{\sum_{i=1}^{N} \frac{mv_i^2}{2}}{N} = \frac{m}{2} \frac{\sum_{i=1}^{N} v_i^2}{N} = \frac{m\overline{v^2}}{2} \qquad (8)$$

here $\overline{v^2} = \sum_{i=1}^{N} v_i^2 \Big/ N$ is **mean-square velocity** of particles (it should be remembered that $\overline{v^2} \neq (\bar{v})^2$). Now using (8) we can rewrite (6) as

$$p = \frac{1}{3} mn\overline{v^2} \qquad (9)$$

How does $\bar{\varepsilon}$ depend on macroscopic parameters of a perfect gas? Let us multiple both parts of (7) by the molar volume V_μ, so we would have

$$pV_\mu = \frac{2}{3} nV_\mu \bar{\varepsilon} = \frac{2}{3} N_A \bar{\varepsilon} \qquad (10)$$

where $nV_\mu = N_A$ is the Avogadro number. Comparing (10) with the equation of state of one mole of a perfect gas

$$pV_\mu = RT \qquad (11)$$

we find that $RT = \frac{2}{3} N_A \bar{\varepsilon}$ and finally we have got

$$\bar{\varepsilon} = \frac{3}{2} \frac{R}{N_A} \cdot T = \frac{3}{2} k \cdot \grave{O} \qquad (12)$$

here in (12) $\frac{R}{N_A} = k$ is the Boltzmann constant, $k = 1{,}38 \cdot 10^{-23}$ *Joule / Kelvin.* It can be considered as the universal gas constant of one molecule.

An important conclusion can be drawn from equation (12) that ***at the state of thermal equilibrium absolute temperature of a perfect gas is proportional to average energy of translational motion of its molecules*****.** (We should recall here that any substance in the state of thermal equilibrium gains the same amount of energy as loses in the same time.)

It turned out that this conclusion is valid not only for a gas, but for a substance in any other aggregate state too, if it is in thermal equilibrium.

Substituting $\bar{\varepsilon}$ from (12) into (7) we can derive

$$p = \frac{2}{3} n\bar{\varepsilon} = \frac{2}{3} n \cdot \frac{3}{2} kT = nkT \qquad (13)$$

From equation (13) it follows that as $n = p/kT$ than *concentrations of one mole of all gases at equal values of pressure and temperature are equal too*. At normal conditions $n = 2{,}69 \cdot 10^{25}\ m^{-3}$, this value is called the **Loschmidt number.**

For a mixture of different gases expression (13) can be rewritten as

$$p = nkT = kT \sum_{i=1}^{l} n_i = n_1 kT + n_2 kT + ... + n_i kT + ...$$

here $p_i = n_i kT$ means the partial pressure of a given gas. It is assumed in the latter formula that a vessel contains l different gases. So we get from the last equation that total pressure of a mixture of gases is equal to sum of all partial pressures of components **(Dalton's law)** or

$$p = \sum_{i=1}^{m} p_i = p_1 + p_2 + ... + p_n \qquad (14)$$

The average velocity of molecules of a perfect gas can be expressed through its macroscopic parameters. From (8) and (12) we have

$$\frac{m\langle v^2 \rangle}{2} = \frac{3}{2} kT$$

(angle brackets denote average value of the second power of velocity), then the root-mean-square velocity of gas molecules is found from (letter μ below denotes the mass of one mole of a gas)

$$\langle v^2 \rangle = \frac{3kT}{m} = \frac{3RT}{mN_A} = \frac{3RT}{\mu} \Rightarrow \sqrt{\langle v^2 \rangle} = \bar{v}_{rms} = \sqrt{\frac{3RT}{\mu}} \qquad (15)$$

3. Internal energy of a perfect gas. Energy distribution on degrees of freedom.

In molecular physics total **internal energy** of a perfect gas is a sum of kinetic energies of all its molecules (molecular interactions are neglected).

In the previously derived Equation (12) for mean kinetic energy of a molecule only kinetic energy of translational motion is taken into account. But together with translation of molecules their rotation and oscillations are possible too. So some amount of energy is stored in these kinds of motion. Therefore we can ask what part of energy corresponds to any kind of motion? To answer this question, the concept of **degrees of freedom** is introduced. Every independent kind of motion of a given particle is called a degree of freedom. For instance, any unfixed body has three translational degrees of freedom as it can move to left or to right, downward or upward, toward the observer of outward of him (her). This motion can take place along three spatial coordinates *X, Y* and *Z* in general case. In addition, a solid can rotate around three nonparallel directions, this motion can be described by three angles α, β and γ. For example, a ball at a horizontal plate has two degrees of translational motion and three rotational degrees, so its position in any moment of time can be given with five coordinates *x, y,* α, β and γ (axis *Z* at *Fig. 2* directed downwards, away from a reader).

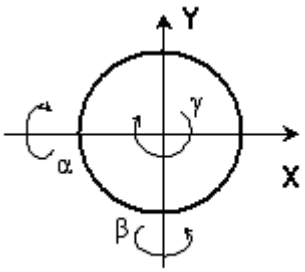


Fig. 2

This example shows that number of possible degrees of freedom of a particle corresponds to the number of independent coordinates needed to describe its state in any moment of time. Apart from translational and rotational motion of objects, oscillations of their parts with respect to each other are possible too. Having been based on the assumption of chaotic character of molecular motion James Clerk Maxwell came to the principle of uniform distribution of molecules on degrees of freedom. It means that equal amount of energy corresponds to each degree of freedom of translational as well as of rotational as well as of oscillatory motion. As energy corresponding to any of three translational degrees of freedom of a molecule obeys the formula (12), to one degree of freedom corresponds one third of that amount of energy

$$\frac{\bar{\varepsilon}}{3} = \frac{kT}{2} \qquad (16)$$

Then kinetic energy of any given molecule is proportional to number of possible degrees of freedom j

$$\bar{\varepsilon} = \frac{jkT}{2} \qquad (17)$$

Atoms in molecules of a perfect gas can be considered as material points therefore any monoatomic gas has only **three** degrees of freedom (*Fig. 3*, a). Kinetic energy of rotational motion $I \cdot \omega^2 / 2$ is infinitesimally small value for a point. Indeed, moment of inertia of a ball-shaped molecule with respect to its axes of rotation is $I = 2mr^2/5$**.** As radius of a point molecule tends to zero, value of its moment of inertia is negligibly small.

One of axes of symmetry of a diatomic molecule coincides with the straight line between atoms (*Fig. 3*, b). We can conclude similarly to the previous reasoning that moment of inertia with respect to this axis is very small, so a diatomic molecule has five degrees of freedom.

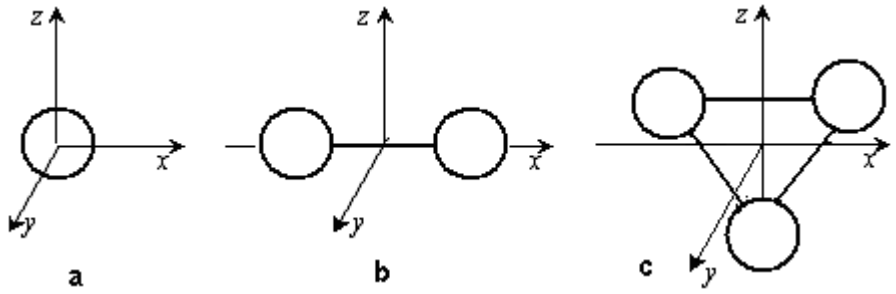


Fig. 3

Molecules consisting of three or more atoms have **three** degrees of freedom of translational **plus three** degrees of freedom of rotational motion.

Thus if number of degrees of freedom of a molecule equals to *j*, then internal energy of one mole of a perfect gas is

$$U_\mu = \bar{\varepsilon} N_A = \frac{jRT}{2} \qquad (18)$$

where j = 3 for monoatomic molecules, j = 5 for diatomic ones, j = 6 for polyatomic molecules.

Experimental data obtained at normal conditions are in good agreement with Equation (18).

It should be added that elastic bounds between atoms in diatomic and polyatomic molecules should be considered as additional oscillatory degrees of freedom. Why don't we take them into account? It had been one of mysteries of classical physics for decades. Only in XX century it was

found from more complicated quantum-mechanical assumptions that energy corresponding to one degree of freedom of oscillatory motion obeys the expression

$$\varepsilon = hf\left(n + \frac{1}{2}\right), \quad n = 0,1,2,3,... \tag{19}$$

where f is the frequency of oscillation, h is the Planck constant ($h = 6{,}63 \cdot 10^{-34}$ $Joule \cdot second$). It turns out that value $n = 0$ corresponds to so-called zero-point oscillations of particles, which take place always, even at temperature of 0 *Kelvin*. To transmit oscillations to the next energy level, corresponding to $n = 1$ a substantial amount of energy is needed, which cannot be gained in thermal motion of molecules at normal conditions. It was counted that, for example, for a hydrogen molecule energy of oscillatory state with $n = 1$ is comparable with kT if temperature is of the order of 6000 *K* (by the way, this is the order of temperature of the Sun's surface). Therefore heating of a gas at conditions close to normal does not lead to excitation of quantum oscillations and they can be not neglected in calculation of internal energy of gases.

4. Heat capacities of gases

If some body (not only gas in a closed vessel) gets some amount of heat, its temperature, in general case, increases. But to reach the same increment of temperature different bodies need different quantity of heat. We say that their heat capacity is different. Heat capacity of some amount of a substance is the ratio of small increment of heat received by that substance to the resulting value of temperature change:

$$C = \frac{dQ}{dT} \tag{20}$$

Generally in theoretic reasoning, we consider heat capacity of a unit mass of a substance, called specific heat, which is usually denoted with the small letter *c*. In the theory of gases it can be more convenient to study molar heat capacity (corresponding to one mole of a gas). It is sometimes denoted as C_μ. To avoid cumbersome notation we will further denote with capital letter *C* molar values of heat capacity, omitting subscript μ. It should be also noted that the latter formula (20) is also valid for specific heat or molar heat capacity.

Heat obtained from external source is used to increase internal energy of a gas (to force its molecules to move faster, to collide more often and so on). It is also known from the equation of state of a perfect gas that the more is its temperature at constant pressure, the more is its volume. But a gas, when expanding, is able to provide useful work like motion of pistons (for example, in steam engines, or internal gas engines). So in definition of heat capacity (20) we can replace small increase of heat by sum of increase of internal energy and of a useful work:

$$C = \frac{dU + pdV}{dT}$$

Internal energy is a function of temperature and volume in general case, it means that

$$dU = \left(\frac{\partial U}{\partial T}\right)_V dT + \left(\frac{\partial U}{\partial V}\right)_T dV$$

and

$$C = \left(\frac{\partial U}{\partial T}\right)_V + \left[p + \left(\frac{\partial U}{\partial V}\right)_T\right]\frac{dV}{dT}$$

But it was found experimentally (Joule and Thomson) that internal energy of a perfect gas at constant temperature does not depend on volume, that is $\left(\partial U / \partial V\right)_T = 0$. So heat capacity obeys the equation:

$$C = \left(\frac{\partial U}{\partial T}\right)_V + p\frac{dV}{dT}$$

Value of specific heat of a gas depends on experimental conditions. If volume of a gas is maintained constant, then its heat capacity at the constant volume is equal to

$$C_V = \left(\frac{\partial U}{\partial T}\right)_V$$

The heat capacity of a perfect gas at the constant pressure C_p is always more than at the constant volume, because derivative dV / dT is

positive according to equation of state of a perfect gas. Using the latter equation, we can write

$$C_p = C_V + R \tag{21}$$

The latter formula is called the **Robert Meyer's equation**.

Using equation (18) obtained in the previous paragraph we can write that for one mole of a perfect gas at constant volume its specific heat is equal to

$$C_V = \left(\frac{\partial U_\mu}{\partial T} \right)_V = \frac{jR}{2} \tag{22}$$

Whereas from (21) we have at constant pressure

$$C_p = \frac{jR}{2} + R \tag{23}$$

It follows from (22)–(23) that molar heat capacities of perfect gases containing different molecules differ. So for a monoatomic gas (j = 3)

$$C_V = \frac{3}{2}R, \quad C_p = \frac{5}{2}R$$

A question: using results (22)–(23) derive expressions for molar heat capacities at constant volume as well as at constant pressure for diatomic and polyatomic gases.

Ratio

$$\gamma = \frac{C_p}{C_V} = \frac{c_p}{c_V} \tag{24}$$

is called the **adiabatic constant** of a gas. It is used in studies of thermodynamic processes, in which a system under consideration does not exchange heat with ambient medium, - adiabatic processes. The most widely spread adiabatic process in gases or liquids is sound propagation. Oscillations of pressure in sound take place so quickly that different parts of a substance do not have time to transmit heat to each other.

From (22)–(24) it follows that $\gamma = 1.67$ for a monoatomic gas, $\gamma = 1.4$ for a diatomic gas and $\gamma = 1.33$ for a polyatomic gas.

5. Maxwell distribution of gas molecules on speeds

Gas molecules move chaotically so value and direction of speed of motion of every molecule often changes after collisions with other molecules and with walls of a vessel. At normal conditions each molecule experiences about 10^9 collisions every second. All directions of motion are equally probable. But what can we say about possibility of different values of speed? In other words this problem can be reformulated so: what number of molecules has absolute values of their speeds in a narrow interval from v to $v + dv$? To solve it, James Clerk Maxwell (1859) supposed that as the total number of molecules was huge, then number of speeds of molecules belonging to a narrow interval should be proportional to the total amount of molecules N as well as to width of the interval dv:

$$dN(v) = f(v) \cdot N \cdot dv \tag{25}$$

here the factor $f(v)$ is called the **distribution function** which determines probability for a molecule to have value of velocity equal to v. Using the theory of probability Maxwell had found that

$$f(v) = 4\pi \left(\frac{m}{2\pi kT} \right)^{3/2} v^2 e^{-\frac{mv^2}{2kT}} \tag{26}$$

Typical curves of Maxwell's distribution function are drawn in *Fig. 4* for two different molecular masses ($m_1 < m_2$). Area under each curve is equal to one.

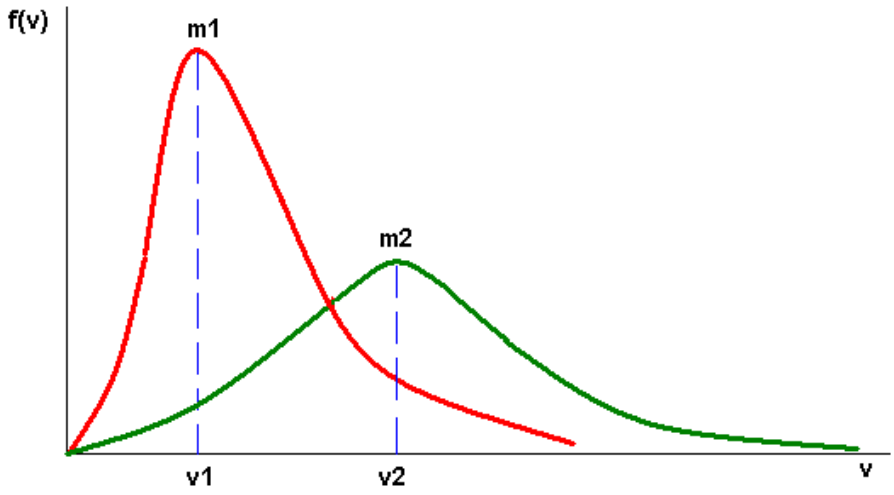


Fig. 4

Value of molecules' speed corresponding to the maximal value of the distribution function is called the **most probable velocity** v_m. After calculating the derivative of *f(v)* and equating it to zero, we can obtain

$$v_m = \sqrt{\frac{2kT}{m}} = \sqrt{\frac{2RT}{\mu}} \tag{27}$$

The most probable velocity is always less than the **mean arithmetic velocity,** which is equal to

$$v_{mean} = \int_0^\infty v f(v) dv = \sqrt{\frac{8kT}{\pi m}} = \sqrt{\frac{8RT}{\mu}} \tag{28}$$

Recalling section 3, we can also introduce the **mean quadratic speed value** v_q, so from equation (16) for a monoatomic gas

$$\bar{\varepsilon} = \frac{m\langle v^2 \rangle}{2} = \frac{3kT}{2} \Rightarrow v_q \equiv \sqrt{\langle v^2 \rangle} = \sqrt{\frac{3kT}{m}} = \sqrt{\frac{3RT}{\mu}} \tag{29}$$

A task: check that $v_m < v_q < v_{mean}$.

Nevertheless, all three values of velocity v_m, v_q, v_{mean} are of the same order and all they can be used to estimate character speeds of gas molecules in thermal motion.

6. Stern's experiment

Many different experiments confirmed Maxwell's theory. One of the most obvious of them had been fulfilled by a German physicist Otto Stern in 1920. Top view drawing of that experimental apparatus is shown at *Fig.* 5. Two rigidly connected coaxial cylinders (*A* and *B* at the Figure) rotated around their common axis with constant angular velocity ω. Thin platinum filament (it is denoted as a black dot *O* in the center of circles *A* and *B*) coated with silver was placed along the axis within cylinders. Strong electric current passed through the filament and heated silver to melting temperature. Evaporated silver atoms behaved like gas molecules at appropriate temperature, they randomly flied away in all directions, colliding each other and accumulating at the walls of cylinders. Some of atoms left inner cylinder through the thin slot and settled at the wall of outer

cylinder. The slower was velocity of an atom the far from the straight line *AC* was it settled (only atoms with infinite speed of motion could be able to reach point *C* when cylinders were rotating). Investigation of relative thickness of silver layer at the wall of the cylinder *B* proved validity of Maxwell's distribution law in this way.

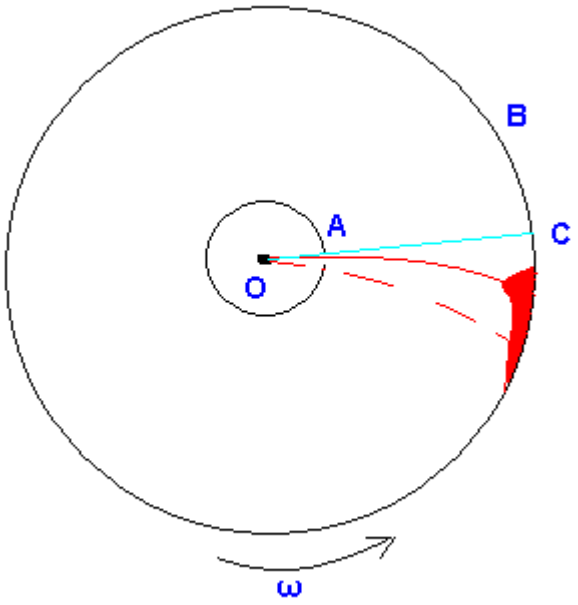


Fig. 5

Calculations using formulas (27)–(29) show that mean speed of gas molecules is of the order of some hundreds of *m/s*: ~ 490 *m/s* for oxygen, ~ 1900 *m/s* for hydrogen.

7. Boltzmann distribution

J.C.Maxwell, when considering distribution of gas molecules, assumed that molecules' concentration is constant in all parts of volume because of their incessant random collisions. In fact, action of all external forces was neglected. But in the Earth's atmosphere all molecules experience action of the gravitational field of our planet. How can it change molecules' distribution? This question was put and answered by Ludwig Boltzmann.

He supposed that because of gravity gas pressure is different at different altitude: the more is vertical coordinate of a point the less is pressure. Then pressure difference between two layers of gas at altitude h and $h + dh$ is very small and proportional dh if dh is small:

$$dp = (p + dp) - p = -\rho g dh$$

sign "-" indicates that pressure diminishes as altitude grows. From equation of state of a perfect gas mass density can be expressed as

$$\rho = \frac{m}{V} = \frac{\mu p}{RT}$$

So we have an ordinary differential equation in p as a function of h:

$$\frac{dp}{p} = -\frac{\mu g}{RT} dh \qquad (30)$$

Its solution looks like

$$p = p_0 \exp\left(-\frac{\mu g h}{RT}\right) \qquad (31)$$

where p_0 is pressure at zero altitude (at the Earth surface, for example). Eq. (31) is called the **barometric height formula**.

We can recall from (13) that $p = nkT$, then barometric height formula transforms into the **Boltzmann distribution** of gas molecules in altitude:

$$n = n_0 \exp\left(-\frac{\mu g h}{RT}\right) \qquad (32)$$

where n_0 is the concentration of a gas at zero level. Index of exponential function in formulas (31)–(32) contains molar mass of a gas μ and it has been confirmed experimentally that in compliance with the Boltzmann distribution concentration of hydrogen decreased with height over Earth's surface sooner than concentration of heavier molecules of nitrogen or oxygen.

8. Transport phenomena in gases

Transport phenomena are called irreversible processes in a thermodynamically non-equilibrium system, which lead to equalizing of previously non-uniform parameters of the system. Concept of transport phenomena embraces **heat conduction, diffusion, viscosity** (or **internal friction**) and **electric conduction**. The latter two phenomena will be considered in future lections.

It had been clear since the middle of XIX century that the reason of all transport phenomena was chaotic thermal motion of particles, from which a given substance is composed – molecules or atoms, or ions. Transport phenomena are rather slow: smell of an odorous substance sprayed in one corner of a room is felt at a distance of some meters not until some seconds although mean speeds of air molecules are of the order of hundreds or even thousands of meters per second (see section 6). It was Clausius who first realized that straightforward motion of gas molecules interrupted with their collisions, after which values and directions of molecular motion substantially changed. Calculations based on the kinetic theory of gases showed that a **mean free path** of gas molecules is of the order of

$$\bar{\lambda} = \frac{1}{\pi d^2 n} \tag{33}$$

where d is the molecular diameter, n is concentration. For simplicity only the case of monomolecular gas is considered here. If a vessel is filled with a mixture of different components, the latter expression becomes more complicated, but the main characteristic features of transport phenomena remain the same.

Diffusion is a process of transport of mass from a point with higher concentration to a point with lower concentration. **Fick equation** states that flux density of molecules (that is number of molecules that intersect a unit area in unit time in the process of diffusion) is proportional to the gradient of concentration:

$$J = -D\frac{dc}{dx}$$

where D is the coefficient of diffusion. So if we consider change of concentration of molecules in some cylinder places along the x-axis between planes x and $x + dx$ (*Fig. 6*) we could easily realize that this change in time is proportional to flux difference at the ends of the cylinder $J(x)$ and $J(x+dx)$:

$$\frac{\partial c}{\partial t} = -\left[J(x+dx) - J(x)\right] \approx -\frac{\partial J}{\partial x} = D\frac{\partial^2 c}{\partial x^2}$$

Equation

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} \tag{34}$$

is called the **differential equation of diffusion**. Coefficient of diffusion is measured in m^2/s in the International System of Units. Using the molecular-theory approach it have been found that

$$D = \frac{1}{3}\bar{v}\bar{\lambda} \tag{35}$$

where $\bar{v}$ is any of average molecular velocities (27)–(29), $\bar{\lambda}$ is the mean free path (33).

Let us consider heat transfer now. Suppose gas in some volume is heated from the left side, so heat flows from left to right (*Fig. 6*). We can mentally choose in the volume a round cylinder with very short height *dx* (this height is greatly magnified for clearness in the *Fig. 6*). Axis of the cylinder coincides with the direction of a coordinate axis *X*.

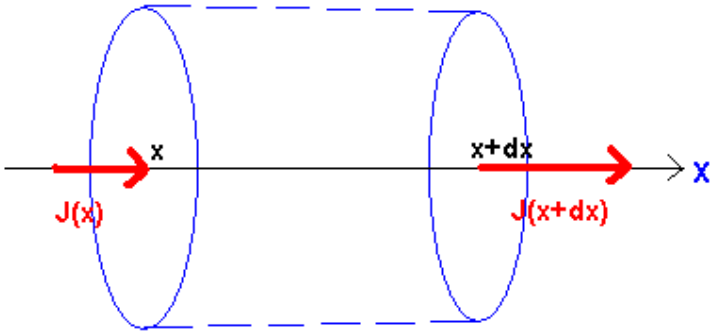


Fig. 6

Some heat source can take place in the cylinder, therefore temperature rises in time and change of heat is proportional to its decrement:

$$dQ = c_V \rho dT$$

In the same time heat can pass only through the circular bases of the cylinder (but not through the lateral walls). Suppose it flows in through the left base and flows out through the right one, like it is drawn at *Fig. 6*. Using notion of the **heat flux density** (heat power per unit area) we can write:

$$J(x) - J(x + dx) \approx -\frac{\partial J}{\partial x} dx$$

where *J(x)* is now the heat flux density at a point *x*.

Combining two previous formulas in one equation, we have the next partial differential equation (in fact, this is equation of conservation of energy in the cylinder):

$$c_V \rho \frac{\partial T}{\partial t} = -\frac{\partial J}{\partial x} \tag{36}$$

It is also known that heat flux takes place only in the case of temperature difference, so

$$J = -\kappa \frac{\partial T}{\partial x}$$

where κ is the **thermal conductivity coefficient**.

Differentiating the latter equation on *x* and substituting the result in Eq. (36) we obtain differential **equation of thermal conductivity** (or **heat conductivity equation**):

$$c_V \rho \frac{\partial T}{\partial t} = \kappa \frac{\partial^2 T}{\partial x^2}$$

or

$$\frac{\partial T}{\partial t} = \chi \frac{\partial^2 T}{\partial x^2} \tag{37}$$

where $\chi = \kappa / c_V \rho$ is the **thermal diffusivity coefficient,** ρ is the mass density. It has been stated in the kinetic theory that

$$\chi = \frac{1}{3} \bar{v} \bar{\lambda} \tag{38}$$

here $\bar{v}$ and λ are the same as in (35). The more widely used the **thermal conductivity coefficient** is equal to

$$\kappa = \frac{1}{3}\bar{v}\bar{\lambda}\rho c_v \qquad (39)$$

It should be emphasized that equations (34) and (39) are identical from the mathematical point of view. Mathematicians had thoroughly studied this kind of equation for more than century, so all numerous solutions of the diffusion equation can be used for the thermal conductivity one and vice versa.

When talking about internal friction the next imaginary experiment is usually considered. There are two flat plates in a viscous medium (gas or liquid). One of the plates is dragged with a constant speed *v* (*Fig. 7* below). Nearest to the moving plate molecules of the liquid

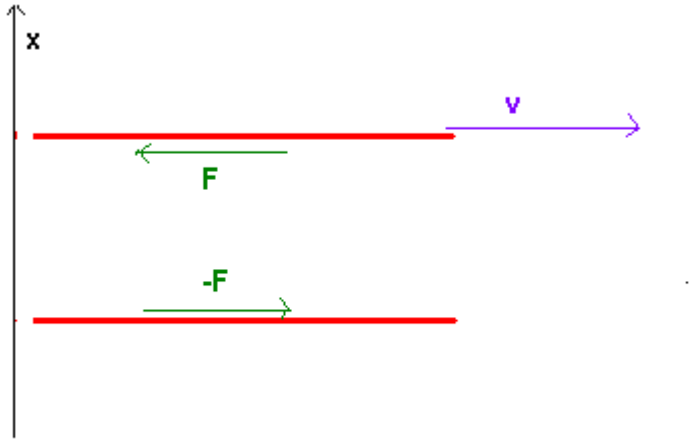


Fig. 7

(their layer has thickness of the order of $\bar{\lambda}$) are carried away in the same direction as the plate with the same speed *v*. That means that some mechanical work is necessary to drag that thin layer of a liquid. Other layers impede this motion that becomes apparent as a force of viscous friction acting in the direction opposite to direction of the plate's speed. Even if distance between plates exceeds $\bar{\lambda}$, a force from the medium acts at the lower (*Fig. 7*) immovable plate in the direction of motion of the upper plate. As Sir Isaac Newton stated, this force is equal to

$$F = \eta S \frac{dv}{dx}$$

where η is coefficient of viscous friction (or viscosity coefficient), *S* is the side area of the plate, axis *x* is directed perpendicularly to plates. Coefficient of viscous friction is measured in $Pa \cdot s$, there is also an off-system unit *poise* (*P*): $1\ Pa \cdot s = 10\ P$. It has been found that in gases

$$\eta = \frac{1}{3}\rho\bar{v}\bar{\lambda} = D\rho \qquad (40)$$

In gases this coefficient is usually grows with the growth of temperature as chaotic collisions become more often and exchange with momentum between molecules accelerates.

9. Real gases. Van der Waals equation.

Until now all our consideration has been in the scope of the main assumptions of the molecular theory (section 1). Equation of state of a perfect gas in that scope cannot explain some evident phenomena like condensation of gases. Therefore now we make a step beyond that scope and consider behavior of real gases.

The main difference between real and perfect gases is that interaction of molecules is taken into consideration in the former case. Molecules of all gases are electrically neutral at normal conditions. Nevertheless, their interaction is of electrostatic origin. Centers of mass of positive and negative charges are in different points in some molecules like those of water. From electrical point of view any water molecule forms a **dipole** as a pair of charges of opposite sign placed at some distance from each other. Theoretical description of electric dipoles will be considered later, but now we can realize how dipole molecules behave at some small distance between them. If that distance is more than molecular diameter, they tend to turn so that opposite charges would be closer to each other, and an attractive Coulomb force appears in this case. If distance between two dipoles-molecules becomes less than their diameter, then electron shells of atoms coincide and forces of electrostatic repulsion between electrons arise. Mathematically those forces of molecular dipole-dipole interaction are described with the help of the **Lenard-Jones potential**, which depends from the distance *r* between the centers of dipoles as

$$U(r) = \frac{a_1}{r^{12}} - \frac{a_2}{r^6} \qquad (41)$$

where a_1 and a_2 are some constant characterizing a couple of molecules. The first term in (41) is responsible for repulsion and the 12^{th} power of its denominator points out a very short radius of action, the second term describes attraction, which becomes apparent at longer distances. Graph of the function (41) is presented in the *Fig. 8*. The point, at which the Lenard-

Jones potential is equal to zero is considered to be of the order of the diameter of a molecule (*d* at *Fig. 8*).

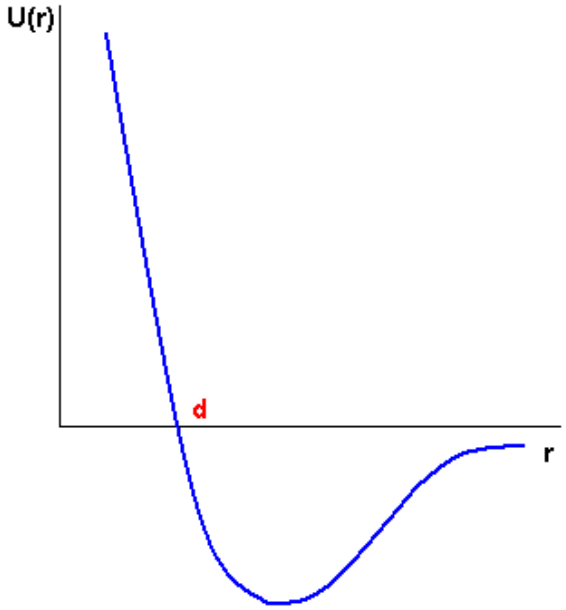


Fig. 8

A force of dipole-dipole interaction at any *r* is the derivative of the potential (41) on the radius taken with opposite sign:

$$F(r) = -\frac{\partial U}{\partial r} = 12\frac{a_1}{r^{13}} - 6\frac{a_2}{r^7} \qquad (42)$$

In 1873 a Dutch physicist Van der Waals based on idea of dipole-dipole interactions proposed the next equation of state of a real gas:

$$\left(p + \nu^2 \frac{a}{V^2}\right)(V - \nu b) = \nu RT \qquad (43)$$

where ν is the number of moles, *a* and *b* are parameters specific for a given gas. The constant *a* is proportional to the attractive force between molecules. The constant *b* is proportional to volume occupied with all gas molecules within a vessel.

Difference between a perfect gas obeying the well-known equation of state (recall the 2nd section of this lecture) and a real gas (equation (43)) is seen from behaviour of isothermal curves – lines presenting dependence of gas pressure on volume at a constant temperature. Isotherms of a real gas

are schematically drawn at Figure 9 for different temperatures (T_1 for the 1st curve is more than T_2 for the second one, analogously $T_2 > T_3$).

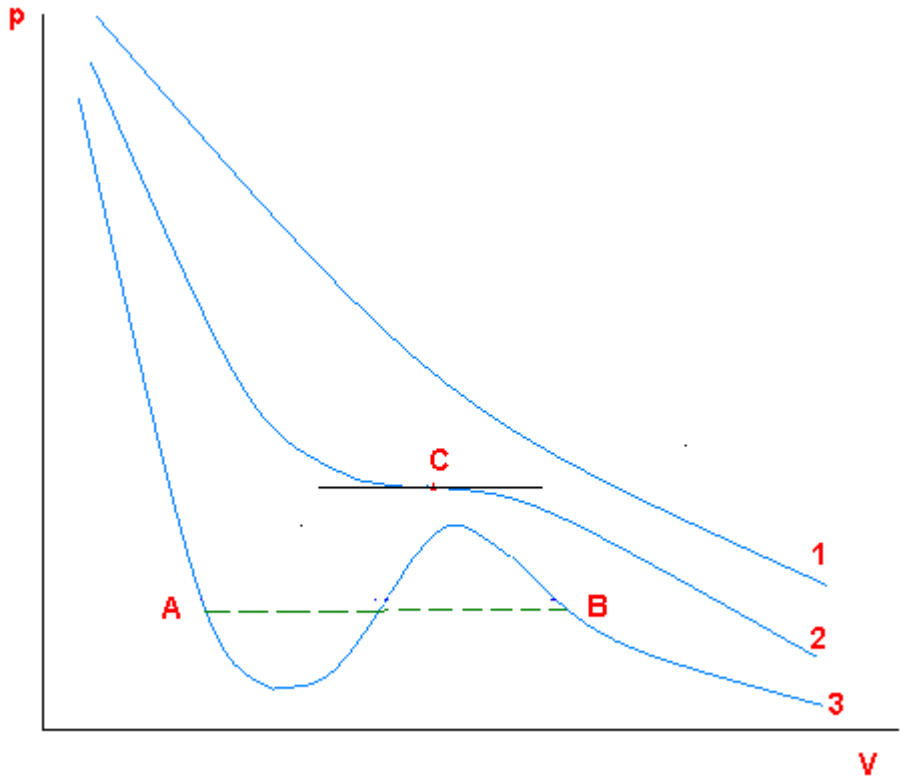


Fig. 9

It is seen that only at high temperatures (1st curve) or for large volumes (right parts of all curves) behavior of isotherms of a real gas resembles hyperbolic dependence character for isotherms of a perfect gas (which obeys Boyle-Mariotte law). The **critical point**, at which tangent to that curve is parallel to the horizontal axis, can be found only at one of isothermal curves. This is the curve 2 and the critical point is denoted with letter *C* at *Fig. 9*.

It has been determined experimentally that pressure of a real gas, which was being compressed at constant not very high temperature obeys Van der Waals Equation (43) only for sufficiently large volumes (the very right part of the 3rd curve at *Fig. 9*). When volume decreases to some value (corresponding to the point *B*), isothermal compression further takes place at constant pressure (along the dashed line *AB*). Only for a very qualitatively cleaned gas a curved line connecting points *B* and *A*, which has one maximal and one minimal values of pressure, can be observed experimentally. This is an example of a metastable process, and a small external influence like vibration can turn such process to a stable one, shown with the dashed line. Usually (in the stable state) as the gas passes from the point *B* to *A* the growing amount of a liquid condenses in the

vessel. Finally, at the point *A* all gas is turned into the liquid phase (phases and phase transitions will be considered in the next lecture). If gas is compressed at higher temperatures, the range of volumes of existence of both phases in a vessel is shortened and at temperature exceeding the critical one a gas cannot be condensed to liquid in any way. The 2nd curve at the *Fig. 9* corresponds to the critical temperature of a given gas. It has been found from Equation (43) that the critical temperature, volume and pressure of a gas obey the next expressions:

$$T_c = \frac{8}{27} \cdot \frac{a}{Rb}; \qquad V_c = 3b; \qquad p_c = \frac{a}{27b^2} \tag{44}$$

A problem. Derive formulas (44) using an assumption that the point *C* in Fig. 9 is the inflection point of the Van der Waals isotherm, that is the next equations are valid in this point (together with (43)):

$$\frac{\partial P}{\partial V} = 0 \qquad \text{and} \qquad \frac{\partial^2 P}{\partial V^2} = 0$$

It should be mentioned, that there are about a dozen of different equations of state of real gases. Some of them contain more than two additional parameters, nevertheless the Van der Waals equation remains the most popular among them because of its relative simplicity and a lucid physical meaning of parameters *a* and *b*.

LECTURE 4. LIQUIDS AND SOLIDS

1. Basic properties of liquids

We had believed for many years that any substance (depending on temperature and pressure) can exist in one of four aggregate states – they are solid state, liquid, gas or plasma. Since 1995 another one aggregate state is added to this list, it is known as the state of **quantum condensate**. Some properties of gases have been considered in the previous lecture. The current one is devoted to molecular properties of liquids and solids.

Liquids have intermediate properties between gases and solids in many respects. Molecular motion in liquids is especially interesting for biophysicists and physicians because biological organisms consist of liquid on more than 50%. Metabolism, supply of cells with oxygen, removal of products of life activity -all these processes provided with movement of molecules, atoms or ions composing liquids.

Different and continuously varying distances between particles is one of the main features of molecular structure of liquids. Only short-range order is observed in location of molecules or atoms. It means that only the nearest neighbors of every particle have definite orientation and position. At distances of the order of 3 – 4 molecular diameters mutual influence of molecules disappears. That peculiarity of liquids provides isotropy of their properties.

Some liquids with elongated or disk-shaped molecules demonstrate an important exception. Ordered orientation of their molecules can maintain at substantial distances at appropriate conditions. These substances are called **liquid crystals** because due to mentioned peculiarity of their structure, anisotropy of mechanical, electrical, optical and other properties also takes place. (**Anisotropy** of some property is its dependence on a direction in space.)

Thermal motion in liquids can be considered as periodic oscillations (with mean frequency of the order of 10^{12} *Hz*) alternating with spontaneous transitions (“leaps”) of particles from point to point. Duration of stay of a molecule at one point is called **relaxation time** and it strongly depends on temperature

$$\tau = \tau_0 \cdot \exp\left(\frac{E_a}{kT}\right)$$

where *k* is the Boltzmann constant, τ_0 is the mean time of a “stable life” at one position, while E_a is activation energy (that is amount of energy which

is need to break the bonds between the given molecule and its neighbours). Average length δ of a leap is of the order of $10^{-8} - 10^{-10}$ *m*, while average velocities of thermal motion of molecules in liquids are of the order of $\bar{v} \sim 10 m/s$ at normal conditions**.**

Non-spherical molecules of liquids can take part in rotational motion too.

Such behavior of molecules explains main property of all liquids, namely their fluidity. Besides that all liquids are practically incompressible like solids. Pure liquids can also resist to mechanic expansion.

2. Transport phenomena in liquids

Inhomogeneity in distribution of either molecules' concentration or of temperature within a volume of a liquid causes transport phenomena like those previously considered for gases in the Lecture 3. Formulas describing diffusion, heat conductivity and viscosity in liquids formally are the same as for gases, but magnitudes and meaning of appropriate coefficients differ strongly.

So diffusion coefficient for chemically uniform liquid obeys the next expression

$$D = \frac{\bar{\delta}^2}{6\tau_0} \cdot \exp\left(-\frac{E_a}{kT}\right)$$

where $\bar{\delta} \sim 10^{-8}$ *m* is the average displacement of a molecule. For example, at $T = 300$ *K* diffusion coefficient for water is of the order 1.5×10^{-9} m^2/s, while for water steam is 2×10^{-5} m^2/s.

Heat transfer in liquids is a process of transfer of energy of vibrational motion but not of translational one as it takes place in gases. Numerical values of heat transfer coefficients for liquids are approximately 100 times more than for gases at same conditions. Those facts can be explained by higher densities of liquids and therefore by much more frequent interactions of molecules.

Viscosity (internal friction) coefficient for liquids is 10^2 - 10^5 times more than for gases and strongly depends on temperature and pressure. For Newtonian (simple) liquids coefficient of viscosity is proportional to the relaxation time and decreases with temperature growth (as opposed to temperature dependence of viscosity coefficients of gases)

$$\eta \sim \tau_0 \cdot \exp\left(\frac{E_a}{kT}\right)$$

It means that fluidity of liquids increases with temperature growth and reaches the maximal value at the temperature of evaporation. So, viscosity coefficient of water at 1 atmosphere and 0^0C is about 1.8×10^{-3} $Pa \cdot s$ while it is 3.4×10^{-4} $Pa \cdot s$ at the same pressure and 90^0C. But if $t = 0^0C$ and pressure 2×10^4 *atmosphere* it increases some million times. As internal friction is one of consequences of interactions of molecules, due to the Bachinsky law (stated experimentally in 1912) viscosity coefficient for liquids is inversely proportional to molar volume V

$$\eta = \frac{C}{V - b}$$

where C is a constant, b is the coefficient form the Van der Waals equation of state of a real gas.

Motion of a liquid as well as of a gas can be laminar (steady, vortex-free) or turbulent (vortical). Mathematical theory of turbulent flow is not developed up to now (and maybe it cannot be developed at all) because of complexity of this phenomenon. A British physicist and engineer Reynolds in XIX century proposed a numerical criterion of definition of a flow kind. Dimensionless **Reynolds number** is a value, which allows compare relative influence of velocity of motion v and of viscosity (relaxation processes)

$$\mathrm{Re} = \frac{\rho \cdot v \cdot d}{\eta}$$

where ρ is the mass density of a liquid (gas), η is its viscosity coefficient, d is a characteristic length of a problem. For circular-section pipes d is the pipe diameter. If an estimated Reynolds number exceeds some experimentally determined value for a given geometry, then liquid flow is turbulent (velocity is too high, so relaxation processes do not have time to prevent formation of vortices). For example, for cylindrical pipes Re is of the order of 2300.

3. Hagen-Poiseuille formula

One of simplest particular cases is viscous liquid's flow through a circular vessel.

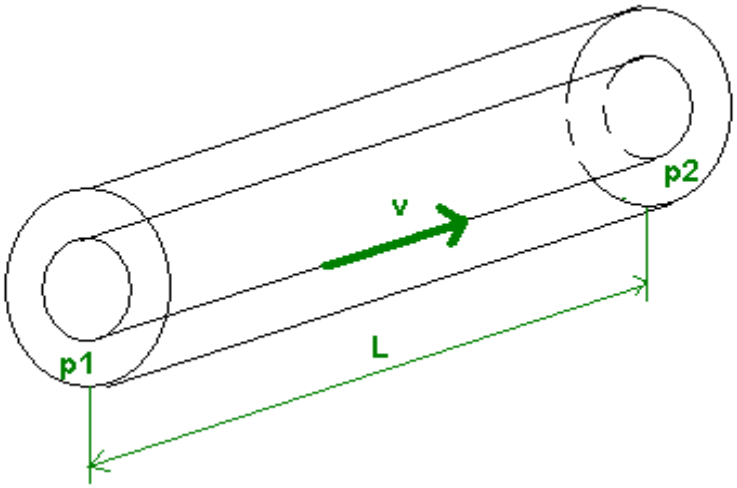


Fig. 1

Let us consider a piece of a pipe of length L (*Fig.* 1) and radius R. Suppose a viscous liquid flows though from left to right and pressure at the left side is equal to p_1, while at the right side is p_2. The flow is laminar therefore we can believe that all moving liquid could be imagined as a set of numerous **tubes of flow** – these are parallel fluxes with non-intersecting volumes. There is no exchange with particles of the liquid between tubes. We can imagine a round cylinder composed from tubes of flow with radius r. Assume that the central axis of this imaginary "small" cylinder coincides with the axis of the whole pipe. In the direction of flow two forces act at the flowing liquid in the small cylinder, they are pressure difference at the ends of the cylinder and the Newton's force of viscous friction (recall the previous lecture) applied to its lateral surface. As the flux takes place with a constant velocity these forces are equal to each other (according to the second Newton law of motion):

$$\pi r^2 (p_1 - p_2) = -2\pi r L \eta \frac{dv}{dr}$$

This equation can be easily integrated giving value of the flux speed as a function of the radius of the "smaller" cylinder:

$$v = -\frac{p_1 - p_2}{4L\eta} r^2 + C$$

where C is an integration constant. We can suppose that near the walls of the pipe (at $r = R$) speed of flow is zero, this assumption allows to determine C, so we get the ultimate expression for the dependence $v(r)$:

$$v = \frac{p_1 - p_2}{4L\eta}\left(R^2 - r^2\right) \qquad (1)$$

It is seen from (1) that viscosity leads to inhomogeneous distribution of liquid speed along the pipe's radius, this speed is equal to null near walls of the pipe and reaches maximal value along the central axis.

Let us now answer the question: what volume of viscous liquid passes through the cross-section of the pipe drawn in *Fig. 1* in a unit time?

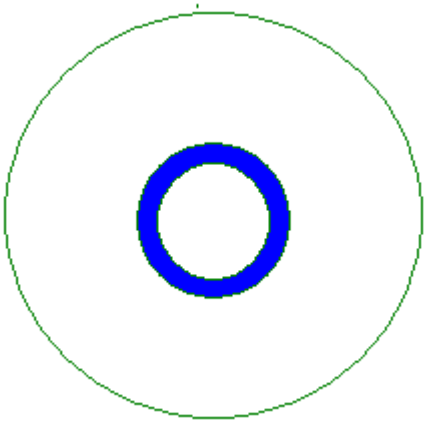


Fig. 2

That cross-sectional area is shown in *Fig. 2*. Outer radius of the pipe is equal to R, as previously. Consider again an imaginary cylinder of radius r being concentric with walls of the pipe. Speed of flow of the liquid along the lateral walls of this "smaller" cylinder is given by (1). But if we can suppose that approximately the same is the speed of flow of all liquid flowing between radii r and $r + dr$. (The liquid located between those radii is shaded in *Fig. 2*). So that volume of the liquid is equal to

$$dV = 2\pi \cdot r \cdot v \cdot dr = \frac{\pi\left(p_1 - p_2\right)}{2L\eta} \cdot \left(R^2 - r^2\right) \cdot r \cdot dr$$

Integrating the left-hand side of this equation from zero to the total volume of the pipe ($V = \pi R^2 L$) and the right-hand side from zero to R, we

obtain a formula named after Hagen and Poiseuille (or simply after **Poiseuille**, though it was really derived by Stokes). It looks like

$$V = \frac{\pi \cdot R^4 \cdot (p_1 - p_2)}{8 \cdot \eta \cdot L} \tag{2}$$

The Poiseuille formula answers the question about amount of a viscous liquid passing through the round pipe with a given length and a given radius. This formula is sometimes used for approximate evaluation of volume speed of liquid flow for pipes of non-circular section too. In derivation of Equation (2) we mentioned a liquid everywhere though it is valid for a gas too.

4. Surface tension

Discontinuity of mass density at a liquid-gas or a liquid-solid interface causes some specific properties of the near-surface liquid's layer. A net force of molecular interaction is zero in the bulk of a liquid, but near-surface molecules undergo net forces directed into the volume perpendicularly to the surface. These forces are called surface tension of liquids. As a result near-surface layer is similar to expanded elastic film and tends to diminish area of liquid surface. A sphere has minimal surface area among all bodies of equal volumes, therefore small liquid drops always tend to maintain spherical shapes.

Force of surface tension is a distributed force because it acts along the line bounded near-surface layer, and its value is proportional to the length of that line

$$F = \alpha \cdot l \tag{3}$$

where α is the **coefficient of surface tension**. This coefficient is measured in *N/m* and it is equal to value of the surface tension force acting at a line of unit length.

From another point of view, if area of liquid's boundary is changed from value *S* to *S + dS* (*Fig. 3*), thus moving a part of boundary of length *L* from a position *x* to *x + dx*, then mechanical work $dA = F \cdot dx = \alpha \cdot l \cdot dx$ is to be performed against the force of surface tension of the liquid. In other words $dA = \alpha \cdot dS$. It means that

$$\alpha = \frac{dA}{dS}$$

so the coefficient of surface tension of a liquid can be considered as mechanical work used for increment of surface area of the liquid, if value of this increment is equal to one. This work is used to enhance the potential energy of molecules of a near-surface layer.

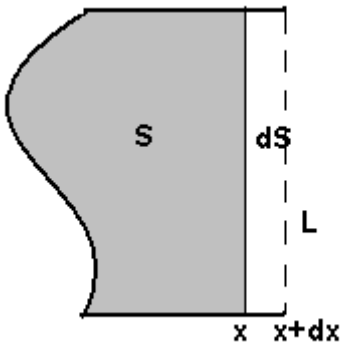


Fig. 3

When some liquid or a solid is dissolved in a solvent, properties of its near-surface layer depend on whether the soluble substance decrease or increase coefficient of surface tension of a solvent. If interaction forces between solvent's molecules are stronger than forces between them and molecules of a soluble substance, than molecules of a solvent mainly migrate from the near-surface layer into the bulk. In this case coefficient of surface tension of solution is less than that of solvent and concentration of molecules of soluble substance is higher in under the surface than in the bulk. Substances diminishing coefficient of surface tension of given solvents are called **surface-active (SAS)** ones. Oil, soap, spirit, fatty acids are surface-active substances with respect to water. SAS badly dissolve in many liquids and it has been found that their molecules are often placed in near-surface layers of solvents in some order. For example, molecules of acetic acid are arranged near water surface so that CO_2H groups are immersed in water while hydrocarbon chains are directed in surrounding air.

There can also exist substances, which increase coefficient of surface tension of solvents, they are called **surface-inactive ones (SIS)**. Salt and sugar are surface-inactive substances with respect to water.

Surface-active substances are used as stabilizers in production of medicines. Salts of higher carboxylic acids (*CH* from 12 to 18) with alkali metals are used as emulsifying agents, stabilizers of ointments and creams. Salts of aluminic, sulphonic, phosphoric acids are used as bactericidal, fungicidal and disinfecting agents. Esters of cucrose and monobasic higher carboxylic acids do not irritate mucous membranes and eyes, therefore they are used in manufacturing of shampoos, soaps, dental pastes as well as in

production of bowels-solvable coats for pills to prevent action of gastric juice.

Molecular interactions leading to phenomenon of **wetting** take place at the interface between two liquids (1 and 2 at *Fig. 4*) and gas (3) or between liquid (2), solid (1) and gas (3 at *Fig. 5*).

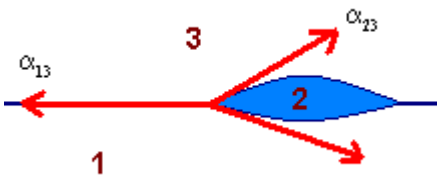


Fig. 4

Drop of a liquid 2 placed at surface of a more consistent liquid 1 is lens-shaped. Three forces are applied to a curved interface between liquids. Each of those forces per unit length is equal to an appropriate coefficient of surface tension: α_{12} is for the interface between both liquids, α_{23} is for the interface between the drop and a gas 3, α_{13} is a coefficient of surface tension at the boundary between the liquid 1 and the gas. First two forces tend to gather drop to sphere while the last one tends to stretch out the drop into a thin layer. If $\alpha_{13} > \alpha_{12} + \alpha_{23}$ then we say that liquid 1 is wetted with a liquid 2.

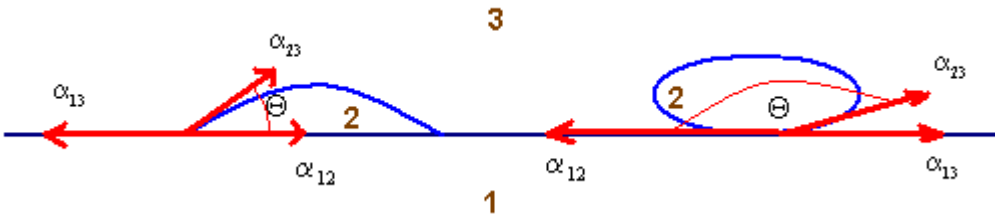


Fig. 5

Analogous effects take place at the gas-liquid-solid interface (*Fig. 5*). Degree of wettability is determined with the **contact angle** Θ. Its apex is at the point of intersection of a plane of a picture and a line belonging to all three media while its sides are formed with tangents to surfaces being in contact. There is high wetting if this angle is acute and low wetting if it is obtuse. It is perfect wetting if $\Theta = 0^0$, and $\Theta = 180^0$ in the case of the perfect non-wetting. Young formula was derived for forces acting at the contact point (*Fig. 5*)

$$\alpha_{13} = \alpha_{12} + \alpha_{23} \cdot \cos\Theta$$

Wetting phenomenon leads to curvature of liquid surface near vessel's walls. That causes additional pressure in the liquid because of tendency of near-surface layer to minimize its area. Laplace was the first who paid attention to that phenomenon and derived the next expression for that additional pressure (sometimes called the Laplace pressure):

$$\Delta p = \alpha\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \tag{4}$$

where R_1 and R_2 are radii of curvature of a liquid surface in two mutually perpendicular directions. In a particular case of a spherical liquid surface, Laplace pressure obeys the simpler equation

$$\Delta p = 2\alpha\frac{1}{R}$$

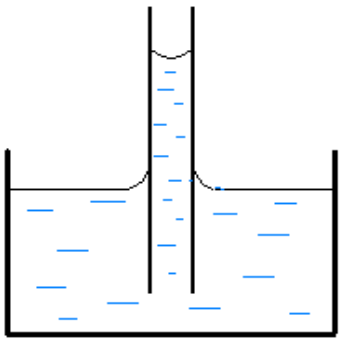


Fig. 6

Applying the Laplace formula (4) to a capillary tube sunk into a vessel with a liquid, which is wetting with respect to a material of the tube (diameter of the tube is overstated at *Fig. 6*), one can find that increase of level of the liquid within capillary is higher than that in the vessel by value

$$h = \frac{2\alpha}{\rho g R}$$

where R the is radius of the tube and α is the coefficient of surface tension of the liquid, ρ is its mass density and g is the gravitational acceleration. This expression can be easily obtained in assumption that gravitational pressure of the liquid in the tube should be compensated with the Laplace pressure in equilibrium.

5. Phases and phase transitions

Phase is a macroscopic physically uniform part of a substance separated from other parts of that substance with an interface, so that the considered phase can be extract from the system mechanically. So it is well known that a vessel, partially filled with water and afterwards closed, always contains some amount of evaporated steam, i.e. the vessel contains two-phase system of a liquid and a gas. If one puts into that vessel a piece of ice, it would be three-phase system until ice melts. Addition of some amount of spirit would not change number of phases because solution of alcohol in water is a uniform liquid. But if, instead of alcohol, to place a drop of mercury into the vessel, one would get a system with two liquid phases, because water and mercury do not mix. Dropping some crystals of salt at the same vessel one could get two solid phases. So a system can contain some different liquid or (and) solid phases but gaseous phase is always one.

Theory of phases mainly considers mechanical and thermal conditions of equilibrium of a system composed of two or more phases. The necessary condition of thermal equilibrium of substances is equality of their temperatures. Parity of pressures is a necessary condition of mechanical equilibrium – for flat interfaces between phases. If boundary of some liquid phase is curved, Laplace pressure of forces of surface tension has to be included into consideration.

In addition, at constant pressure and temperature two adjoining phases can turn into each other or – in other words – to experience phase transitions. So in a closed partially filled vessel with water liquid and gas continuously exchange with some amount of molecules. Therefore state of equilibrium between phases implies another one condition: balance with respect to **phase transitions**.

Changes of aggregate states of substances are evident examples of phase transitions. These are evaporation and condensation, melting and solidification. Less evident is a phase transition called sublimation or transition of a solid into gaseous phase immediately. House-ladies use this phenomenon when drying linen in open air on a frosty day.

As constancy of temperature and pressure play important role in phase transitions, these processes are usually considered at the *TP*-plane (*Fig.* 7).

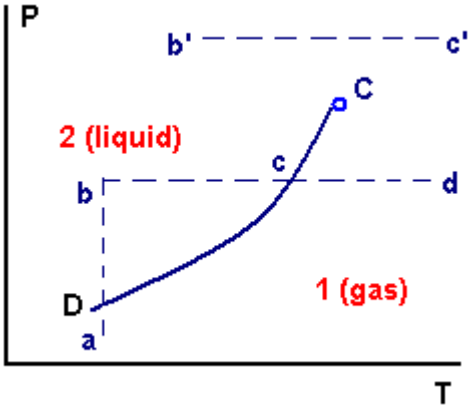


Fig. 7

We now use *TP*-plane for illustration of processes of evaporation and condensation although the same consideration can be applied for any other couple of phase transitions. Thus, any point at the graph of in *Fig.* 7 denotes one-phase state of a substance, for instance points *a* and *d* correspond to a gas, while point *b* describes a liquid. Only points at curve line *CD* belong to both phases. Every point at this curve corresponds to liquid, to gas and to the mixture of both phases. Curve *CD* is called the **evaporation curve** and algebraically it represents the dependence $P = P(T)$. It is clear that motion from the state *a* to *b* describes the process of isothermal condensation, for example, of water steam into liquid water with increase of pressure. Similarly, movement from *b* to *d* describes isobaric (taking place at constant pressure) evaporation of water at rise of temperature. *C* is the **critical point**, it has been mentioned in the Lecture 3 that at temperatures exceeding the critical point a gas under consideration cannot be condensed with any increase of pressure. So at *Fig.* 7 an isobaric process *b'* - *c'* does not need any energy consumption, substance remains physically uniform and its properties change gradually. It means that phase transition does not take place in such process. (It is interesting to mention that at critical temperature coefficient of surface tension of a liquid becomes zero.)

Slope of evaporation curve is given with the **Clausius-Clapeyron equation**:

$$\frac{dP}{dT} = \frac{q}{T \cdot (v_1 - v_2)} \tag{5}$$

where q is the specific heat of evaporation, v_1 is the specific volume (that is inverse mass density) of a gaseous phase, v_2 is the specific volume of a liquid.

Three phases of chemically uniform substance can also stay in equilibrium. This case is depicted in *Fig. 8* :

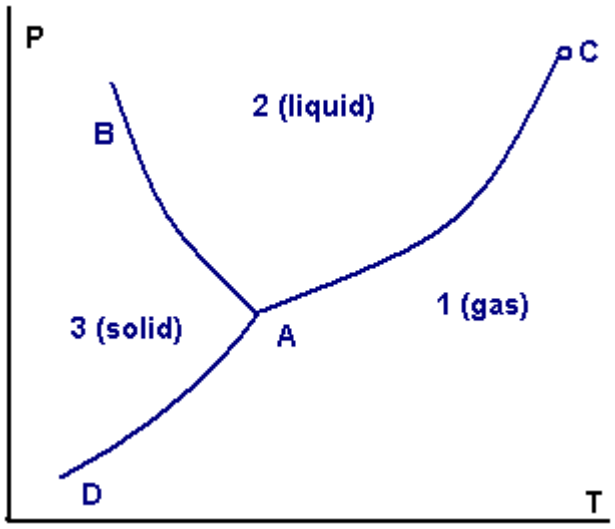


Fig. 8

The **melting curve** *AB* crosses the evaporation curve *AC* at the point *A*, which is called the **triple point**. Three phases can be in equilibrium only in one (triple) point. Of course, everybody can conceive of situations in which three phases of water exist at a range of temperatures, but such states are in general unstable so that at some temperatures, for example, melting prevails while at some lower ones crystallization dominates. Average number of molecules leaving some phase is not equal to average number of molecules replenishing that phase in non-equilibrium state. Slope of the **sublimation curve** *DA* is steeper at the triple point *A* than that of the evaporation curve *AC* because according to Clausius-Clapeyron equation:

$$\frac{dP_{AD}}{dT} = \frac{q_{AD}}{T \cdot (v_1 - v_3)} \quad \text{and} \quad \frac{dP_{AC}}{dT} = \frac{q_{AC}}{T \cdot (v_1 - v_2)}$$

denominators of right-hand sides of these expressions are approximately equal to each other in the triple point, but under the first law of thermodynamics (conservation of energy in thermal processes) $q_{AD} = q_{AB} + q_{AC}$ in the triple point, so $dP_{AD}/dT > dP_{AC}/dT$. Triple point of water at normal pressure is 0.008 ^{o}C over the melting point. *TP*-plane with the triple point and denoted three curves is called the **diagram of state** of a given substance. It is proven that more than three phases of a chemically uniform substance cannot exist together at any pressure and temperature, it means that more three curves cannot intersect at a *TP*-diagram (**Gibbs rule**).

If some substance can exist in more than three phases, its *TP*-plane is separated into some domains. Simplified diagram of state of sulfur is drawn in *Fig. 9*. Solid sulfur can exist as in form of a monoclinic as well as of an orthorhombic crystal (see below the crystal classification), therefore in *Fig. 9* there are three triple points (*S*, *T* and *L*).

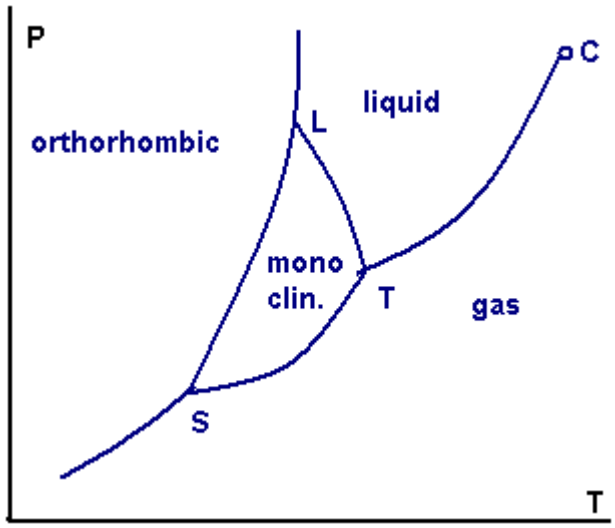


Fig. 9

All considered until now phase transitions have run with some release or absorption of heat, they are called the **first-order phase transitions**. There are also possible other transformations of substances, such as transitions of ferromagnets into the paramagnetic state, transitions of some metals and alloys into the superconducting state at low temperatures, transition of *He* from the liquid to the superfluid phase. They take place without heat exchange and change of volume, but with step-wise change of specific heat and coefficient of thermal expansion. This kind of processes are called the **second-order phase transitions**. Clausius-Clapeyron equation is not valid for the latter kind of phase transformations.

6. Solids

Solids are substances characterized by stability of their shape and volume because of strong interaction forces acting between particles. Thermal movement of particles, which solids are composed from, is confined to small oscillations near positions of equilibrium. There are crystalline as well as amorphous solids.

Amorphous solids as on their main properties and structure can be considered as supercooled liquids with very high viscosity coefficients.

Only the short-range order can be found in location of their particles. Some amorphous solids (such as waxes, resins) are unable to crystallize at all. Other form crystalline lattices under slow cooling, but in opposite case high viscosity prevents ordering of particles' location. Such substances solidify faster than crystallize, they are called vitreous materials (glasses, ice). Sometimes crystallization process in vitreous substances can continue much later after solidification (glass turbidity).

Solid organic substances are also amorphous, such as rubber, wood, leather, plastics, woolen, cotton, silk fibers and so on.

Amorphous solids do not have temperature of melting (solidification). Gradual transition of amorphous substances from the solid to the liquid phase (and backwards) means that there is no specific heat of fusion of those solids.

When crystalline solid reaches the melting temperature, both solid and liquid phases exist simultaneously. Transition of a molecule from the solid to the liquid phase means that it acquires additional energy in translational motion. So unit of mass of a liquid substance has more internal energy than of the same substance in a solid phase. In addition, distances between particles change. Therefore, amount of heat needed for melting of the unit mass of a solid is equal to

$$q_{melt} = \left(u_{liq} - u_{sol}\right) + p \cdot \left(V_{liq} - V_{sol}\right)$$

where q_{melt} is the specific heat of fusion (melting), $\left(u_{liq} - u_{sol}\right)$ is the difference between internal energies in liquid and solid phases respectively, p is an external pressure and $\left(V_{liq} - V_{sol}\right)$ is a difference between volumes in different phases. As the melting temperature depends on pressure, so that from Clausius-Clapeyron equation

$$\frac{dT}{dP} \sim \frac{V_{liq} - V_{sol}}{q_{melt}}$$

It follows from the formula above that if volume of a liquid phase of a substance is more than that of solid phase then growth of pressure increases melting temperature and decreases in opposite case (water, bismuth).

As there is no fusion temperature for any amorphous solid, heating leads to gradual increase of velocity of thermal motion and to diminishing of viscosity coefficient.

The main characteristic property of a crystalline structure is spatial periodicity of internal electric fields and repeatability of location of

particles (**far-range order** of atoms, molecules or ions). Those particles alternate according to definite succession along straight lines, which are called **nodal lines**. Two intersecting nodal lines form a set of identical parallelograms at any section of a crystal with a plane. In 3D space any three sets of nonparallel node lines form a set of identical parallelepipeds, of which a **crystal lattice** consists. A point of intersection of nodal lines is called a **nodal point** (or a **node**) of a crystal. Distance between neighboring nodes at some direction is called **period or translation** of a lattice. A parallelepiped drawn at three nonparallel translations as at sides is called a **unit cell** of a lattice. Symmetry of particles' location with respect to some directions or planes is an important property of a crystal lattice.

Crystal solids can be of two main kinds: single crystals and poly-crystals. A common crystalline lattice takes place in any single crystal. Angles between faces of single crystals of the same symmetry are always the same. An important property of single crystals is anisotropy of their mechanic, thermal, electric, optical and other properties.

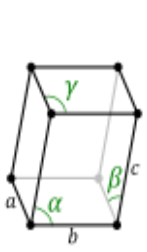
Single crystals can be often found in nature. For example, many minerals (emerald, rock crystal, ruby) are single crystals. Nowadays, many single crystals are also grown from solutions or melted artificially for industrial use – they are ruby, germanium, crystalline silicon, gallium arsenide and many others.

The same chemical element can form different crystalline lattices, this phenomenon is called **polymorphism**. For example, carbon can exist as graphite, graphen and diamond, water ice can be frozen in nature in two crystalline modifications, but 17 more modifications of ice were grown in laboratory environment. Iron is crystallized in $\alpha-$iron, $\beta-$iron, $\delta-$iron and $\gamma-$iron phases.

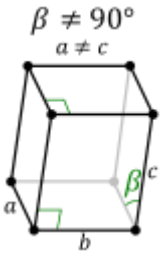
Crystallography is a branch of physics, studying crystals. It numbers 7 different crystalline systems, they are: **triclinic, monoclinic, orthorhombic, rhombohedral, hexagonal, tetragonal and cubic**. Relations for lengths of sides as well as for angles between them are presented in the next table for unit crystalline cells belonging to all crystal systems (compare with figures below):

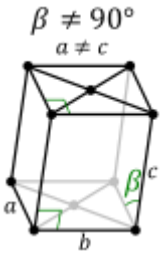
crystallographic system	***lengths of unit cells***	***angles***
triclinic	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma$
monoclinic	$a \neq b \neq c$	$\alpha = \gamma = 90^o \neq \beta$
orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90^o$
``trigonal or rhombohedral	$a = b = c$	$\alpha = \beta = \gamma < 120^o \neq 90^o$
hexagonal	$a = b \neq c$	$\alpha = \beta = 90^o, \gamma = 120^o$
tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90^o$
cubic	$a = b = c$	$\alpha = \beta = \gamma = 90^o$

14 possible elementary **Bravais lattices** belong to those 7 crystalline systems. Those lattices are shown at figures below (names of crystalline systems are in brackets):



triclinic lattice

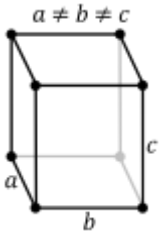




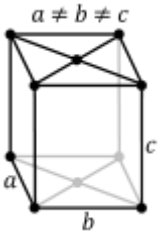
monoclinic

monoclinic base centered

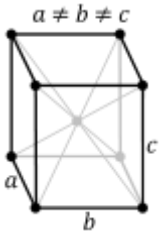
(triclinic system) (monoclinic system)



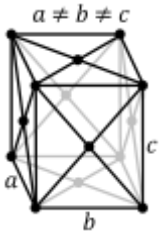
orthorhombic



orthorhombic base centered

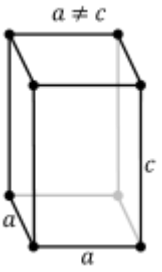


orthorhombic body centered

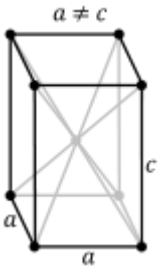


orthorhombic face centered

(orthorhombic system)

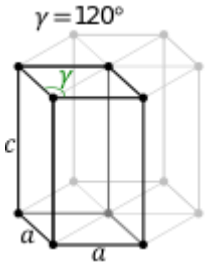


tetragonal



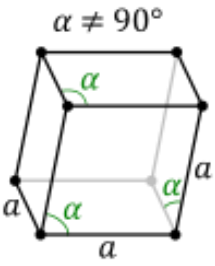
tetragonal body centered

(tetragonal system)



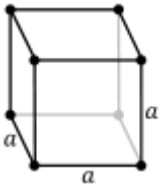
hexagonal

(hexagonal system)

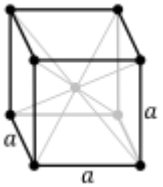


rhombohedral

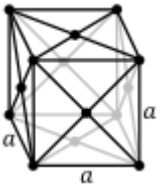
(trigonal or rhombohedral system)



cubic



cubic body centered



cubic face centered

(cubic system)

Fig. 10

Poly-crystalline solids consist of a great number of randomly oriented small single crystals. Therefore regular faceting and anisotropy of physical properties are not typical features of poly-crystals. That is the result of crystallization process: as solidification conditions had been achieved, centers of crystallization appeared in many points of the liquid phase of the substance. Growing crystals are placed and oriented randomly with respect to each other in such process. So as a result of crystallization a solid conglomerate of small chaotically oriented crystals is obtained.

7. Thermal properties of solids

As particles in solids interact strongly with each other, their thermal motion is restricted with oscillations around nods of crystalline lattices (or some points of equilibrium for amorphous materials). Amplitudes of such oscillations are as small as 10^{-11} *m* at normal conditions, or usually 5-7% of a lattice period. Forms of such oscillations are very complicated because it is a result of forces of interaction of an oscillating particle with all its neighbors.

Rise of temperature increases energy of oscillatory motion of particles. That leads to nonlinear increase of amplitude of oscillations and causes thermal expansion of solids which obeys the next dependence

$$l = l_0 \cdot (1 + \alpha \cdot t)$$

where l and l_0 are lengths of a solid spacemen at a given temperature and at 0^oC respectively, α is the **linear expansion factor**. Usually α is of the order of 10^{-6} - 10^{-5} K^{-1}. As a result of thermal expansion volume of a solid increases too

$$V = V_0 \cdot (1 + \beta \cdot t)$$

where β is the **volume expansion factor**. For isotropic solids $\beta = 3\alpha$. As single crystals are anisotropic substances, they usually have three different values of linear coefficient of thermal expansion (for three special dimensions).

It is believed from semi-classical point of view that each particle in each node of a solid has three degrees of freedom of oscillatory motion. As any object involved in oscillations has some amount of kinetic energy as well as some amount of potential energy, it is believed that total energy per a degree of freedom is equal to $\varepsilon = kT$ (that is 2 time more than the energy

per degree of freedom of a molecule of a gas). Thus internal energy of one mole of a solid is equal to

$$U_{in} = 3N_A kT = 3RT$$

So molar specific heat capacity of a solid obeys the next expression

$$C_\mu = 3R$$

or, in other words, the molar specific heat of all chemically uniform crystals is the same and does not depend on temperature. At room temperature this law, named after **Dulong and Petit**, is well satisfied, but as it had been found experimentally it violated at low temperatures. Explanations of reasons of violation of Dulong-Petit law were proposed in quantum theory of heat capacity developed by Einstein and Debye.

Real crystals have some disturbances in their structure called crystals' defects, such as nodes' vacancies (**Schottky defects**), location of particles in space between nodes (**Frenkel defects**), interstitial impurities (alien atoms), edge or screw dislocations (violations of formation of atomic planes) and some othes. Presence of crystalline defects can significantly vary physical properties of crystals, such as thermal and electric conductivity, optical properties.

LECTURE 5. ELECTRIC FIELD. DIRECT ELECTRIC CURRENT.

1. Electric charges and fields

Even ancient Greeks had known that rubbing of some substances like a piece of amber with a textile imparts to amber new properties. On the one hand, it gained ability to attract some small objects, on another one, two rubbed peaces of amber repulsed each other. Now we can explain those phenomena with existence of tiny **electric charges**, which attract each other if they have opposite signs or repulse in the case of the sign of charge. According to the modern point of view formulated by James Clerk Maxwell every electric charge is surrounded with invisible matter called **electric field**, so any change of a position, or speed, or value of the charge is transmitted to other charges by perturbation of electric field, which is known as **electromagnetic wave**.

A French physicist Charles-Augustin de Coulomb stated the law for a force of interaction of two electric charges (1785):

$$F = k\frac{qQ}{r^2} \tag{1}$$

where k is a constant depending on units, q and Q are values of interacting charges, r is a distance between them. Value of electric charge is measured in *Coulomb* (abbreviation *C*) in the International System of Units. Traditionally constant of proportionality was considered to be equal to

$$k = \frac{1}{4\pi\varepsilon_0} \equiv \frac{1}{4\pi \cdot 8.85 \cdot 10^{-12}} \frac{kg \cdot m^2}{C^2}$$

$\varepsilon_0 \approx 8.85 \cdot 10^{-12}\, F/m$ is called the “dielectric constant of vacuum” or “dielectric permittivity of vacuum”, though since 2019 k is considered to be $k \approx 8.99 \cdot 10^9\, kg \cdot m^2 / C^2$.

We can interpret formula (1) as a force acting at a charge q from a another one charge Q (and, according to the third Newton’s law of motion, vice versa). But in the case of many charges, Q can be considered as a sum of all charges placed at the distance r from the charge q. It means that we can characterize an electric field surrounding the charge q with value

$$E = k\frac{Q}{r^2} \tag{2}$$

E is called **intensity of electric field** acting at the charge q in a point of its location. Intensity of electric field in the SI system is measured in $Volt/meter$ (V/m). Strictly speaking, intensity of electric field (as well as the Coulomb force too) is a vector quantity, it is directed from a positive to a negative charge. Therefore the joint action of some amount of charges at the charge q is a vector sum of all forces, which can even compensate each other in some cases.

Michael Faraday proposed to depict electric field with the **force lines** or field lines (see *Fig.1* below). Those lines begin at positive charges and end at negative ones. Vector of intensity of electric field is always tangent to appropriate force line at a given point. The more dense is distribution of field lines at some part of space the stronger is electric field there (so the length of a vector of field intensity is more at a point A of *Fig. 1* than at a point B). You should remember that force lines do not exist in a real world, these are only useful illustrative technique.

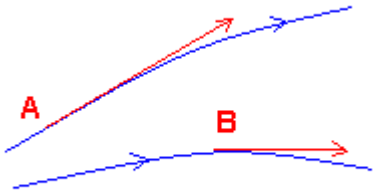


Fig. 1

2. Electric potential

If a charge changes its position at a small distance dx under the action of a Coulomb force F then mechanical work produced by an electric field is equal to:

$$dA = F \cdot dx = qE \cdot dx \qquad (3)$$

Work of electrostatic forces used for motion of that charge between rather distant points A and B is calculated the next integral

$$A = \int_{x_A}^{x_B} F(x) \cdot dx$$

It had been found that value of this integral does not depend on a path along which the charge was really moved, but only on coordinates in the initial and the final points. So this work can be written as a product of the charge value and increment of some parameter characterizing electric field in a given point and called **electric potential** (usually this product is spelt with minus):

$$A = -q(\varphi_B - \varphi_A) \tag{4}$$

In fact, equation (4) is a consequence of inverse square dependence of intensity of the electrostatic field on distance ($\sim 1/r^2$) to the charge, which creates this field. It can be proven that for any other dependence (4) would be incorrect.

Electric potential created with some point charge Q at a distance r from it is equal to

$$\varphi = k\frac{Q}{r} \tag{5}$$

Contrary to the intensity of electric field the potential is a scalar value, it does not have any direction. It is measured in *Volts* (*V*) in the International System of Units due to reasons, which will be clarified later. Graphically electric potential is depicted with equipotential curves, so that at any line value of the potential is constant. It can be shown from formulas (3) and (5) that connection between concepts of intensity of electric field and electric potential obeys the formula

$$E = -\frac{d\varphi}{dx} \tag{6}$$

or, in other words, it is said that intensity is a gradient (directional derivative) of electric potential. That leads to the rule that graphically equipotential lines are perpendicular to field lines. Indeed, in the opposite case some mechanical work would be fulfilled in motion of a charge from one point of an equipotential line to another one. But according to (5) any work of electrostatic force on motion of a charge along an equipotential line is equal to zero.

As electric field provides some mechanical work on motion of an electric charge form point to point, it has some energy. Volume density (value of energy in a unit volume) of energy of electric field at some point is equal to

$$w = \frac{\varepsilon_0 E^2}{2}$$

where E is intensity of the field in that point.

3. Electric dipole

Electric dipole is a system of two electric charges of the same absolute values and with opposite signs placed rather close from each other. Its main characteristic is the **dipole moment**, which is equal to

$$p = q \cdot l \tag{7}$$

where q is charge, l is distance between charges. Dipole moment is also a vector value, it is directed form the negative to the positive charge.

Electric potential generated by a dipole at some point A placed at a distance r much exceeding that between the charges of the dipole can be calculated as algebraic sum of potential created with the positive as well as with the negative poles:

$$\varphi(A) = \frac{q}{4\pi\varepsilon_0 r_+} - \frac{q}{4\pi\varepsilon_0 r_-} = \frac{q(r_- - r_+)}{4\pi\varepsilon_0 r_+ r_-}$$

where r_+ is a distance from the positive charge of the dipole to the point A and r_- is a distance from the negative charge of the dipole to the same point. If $(r_- - r_+) << r_-, r_+$, we can write in denominator $r_+ \approx r_- \approx r$

$$\varphi(A) \approx \frac{q(r_- - r_+)}{4\pi\varepsilon_0 r^2}$$

However, in enumerators of the latter two formulas we cannot write $r_+ \approx r_- \approx r$, because it intuitively clear that the closer is some charge to the point A the more is potential, which it creates, and if the point A is at the same distance from both dipole charges, electric potential generated by this dipole at A would be equal to zero. Electric potential of the same value and opposite sign would compensate each other at A. It means that we need to count this small difference in distances from the dipole charges to the point A. It is seen from the *Fig. 2* that this difference is equal to

$$r_- - r_+ = l \cdot \cos\alpha$$

where α is the angle between axis of a dipole and a straight line directed from its center to the point *A*.

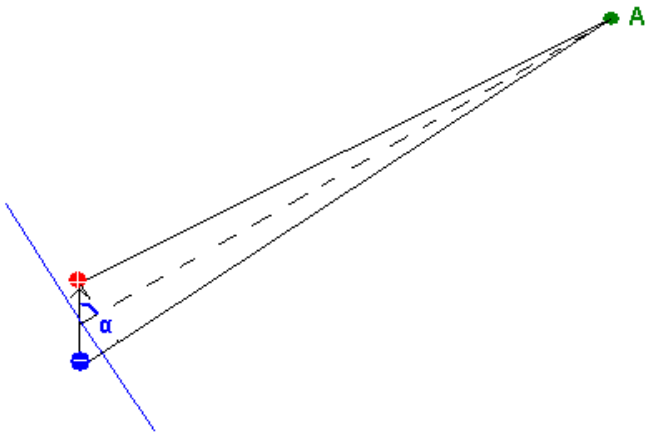


Fig. 2

So we obtain the next expression for electric potential generated by a dipole at some rather distant point:

$$\varphi(A) \approx \frac{ql \cdot \cos\alpha}{4\pi\varepsilon_0 r^2} = \frac{p \cdot \cos\alpha}{4\pi\varepsilon_0 r^2} \qquad (8)$$

Formula for intensity of electric field created by the same dipole in the point *A* is derived in a more complicated manner, so we just mention it here:

$$E(A) \approx \frac{p}{4\pi\varepsilon_0 r^3}\sqrt{1+3\cos^2\alpha} \qquad (9)$$

Value of dipole moment is measured in $Coulomb \cdot meter\,(C \cdot m)$ in the SI system of units, though in atomic physics and chemistry, where distances are of the order of *nanometers* or even *Angströms*, more popular is off-system unit *debay* (*D*). $1D \approx 3.33 \cdot 10^{-30}\ C \cdot m$. You could admit that it is easier to express the dipole moment, for instance, of a water molecule as 1.8 *D*, than in $C \cdot m$. (***A question*** : convert value of the dipole moment of a water molecule into SI units.)

Concept of a dipole is very important in physics and chemistry. We mentioned it in the 3rd Lecture when considering real gases. Forces of

molecular interaction in liquids are also of dipole origin. But from the first glance it could be surprising, because every dipole is electrically neutral. How could neutral object act with electric forces?

Suppose some dipole is placed in a uniform electric field (*Fig. 3*). Uniform, as usual, means that its value is the same in any point. Wherefore filed curves are drawn as equidistant straight lines at *Fig. 3*. This field attracts negative charge to left, but positive one – in opposite direction.

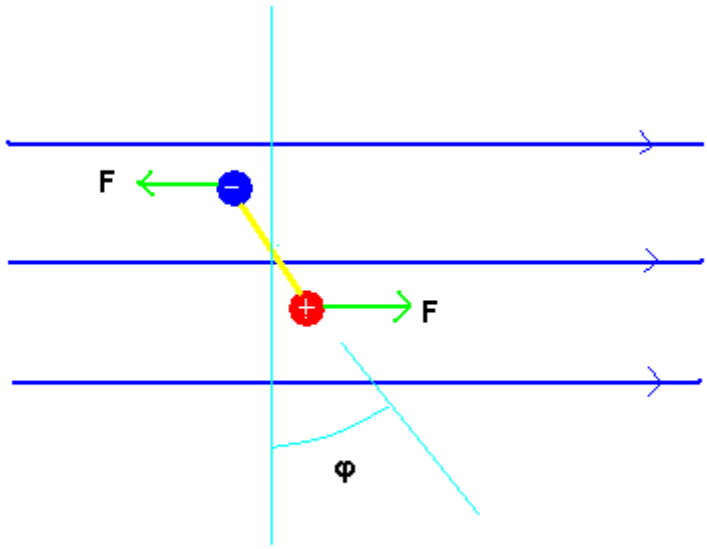


Fig. 3

It means that the dipole is rotated with a torque equal to

$$M = F \cdot \frac{p}{2} \cdot \cos\varphi$$

because only the component of a force perpendicular to the axis of rotation makes a contribution into the torque. So a dipole rotates in a uniform electric field, tending to align its dipole moment along force lines of the field.

Suppose a dipole is placed in a nonuniform electric field, then opposite forces act at different charges, but they don't compensate each other and as a result dipole moves in the direction of the stronger field (*Fig. 4*). At the figure below the electrostatic force acting at the positive charge is denoted with the capital letter because it is more than that acting at the negative charge.

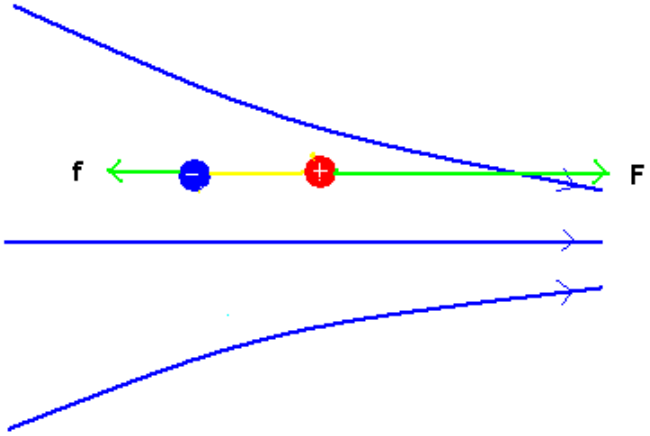


Fig. 4

Mechanical force acting at the dipole in this case is equal to

$$F - f = q \cdot E_+ - q \cdot E_- \approx p \cdot \frac{dE}{dx}$$

This example explains the origin of attractive forces between dipole molecules in a substance when distance between molecules exceeds their sizes. On the contrary, repulsion forces act at distances compared with molecules' sizes, because influence of change of electronic shells of neighbor molecules is to be taken into account in this case.

4. Electric fields in dielectrics and conductors

Dielectrics (insulators) are substances, which cannot transmit electric current (a flux of charged particles), because they do not contain free charge carriers. All gases at normal conditions, many liquids (like water, alcohol, organic oils), some solids (like diamond, kitchen salt) are dielectrics. Molecules of dielectrics in presence of electric field are dipoles. In some dielectrics molecules have dipole moment without applied fields, in others dipole moment is induced by the external field (positive and negative charges drift from each other until their mutual attraction stops their separation). We know that in an external electric field dipoles rotate or move, so their rearrangement diminishes applied field in the bulk of a dielectric. Value of this decrease of field's intensity is given with **dielectric permittivity** ε. Therefore all formulas for Coulomb force (1), intensity of electric field (2),(9) and electric potential (5),(8) written for dielectics should contain ε. For example, expression for Coulomb force is changed to

$$F = k\frac{qQ}{\varepsilon r^2} = \frac{1}{4\pi\varepsilon_0}\cdot\frac{qQ}{\varepsilon r^2}$$

Dielectric permittivity is a dimensionless value.

Quite different situation takes place in conductors. They contain some amount of free charge carriers. In liquids (like solutions of salts) these are ions of opposite sign, in metals (including mercury) these are electrons of conductivity (electrons of outer shells). Electrons of conductivity from classical (non-quantum) point of view can be considered as gas of negative particles, which can freely propagate between ions of a crystalline lattice, like molecules of air in a room. Because concentration of free electrons is very large in typical metals (of the order of Avogadro number) appearance of an external electric field immediately changes distribution of electrons in a metal, so that electric field within it is always equal to zero. Numerical estimates show that very small percentage of free electrons in a typical metal is enough to eliminate an electric field in the interior (or to *screen* the external field as it is said sometimes).

Surface of any metal placed in an electric field becomes the equipotential one, because if it was a potential difference between two points of such surface then a flux of free electrons equalize potentials in those points immediately. As we know that equipotential curves are always perpendicular to force lines of electric field, it means that intensity of a static electric filed is always directed normally to metallic surfaces.

5. Direct electric current in metals

Electric current is a directional motion of electric charges under influence of applied electric field. If the field is constant in time that is **direct current**. In metals free electrons, when conveying electric current, randomly change value and direction of their speed because of interaction with each other and with positive ions of a crystalline lattice. In not very strong electric field we can suppose that every electron has average speed proportional to intensity of applied electric field.

$$\bar{v} = \mu \cdot E$$

where the coefficient of proportionality μ is called **mobility**. Speed of flow of electric charge is called the **intensity (or strength) of electric current.** It can be determined as flux of electric charge or, in other words, product of charge of one electron and number of electrons passing through some area *S* in a unit time:

$$I = \frac{dQ}{dt} = q \cdot n \cdot S \cdot \bar{v}$$

n is the volume concentration of electrons of conductivity. It is seen from the latter equation that flux density of electric current (or **current density**) in a metallic conductor is equal to

$$j = \frac{I}{S} = q \cdot n \cdot \bar{v} = q \cdot n \cdot \mu \cdot E = \lambda \cdot E \qquad (10)$$

Proportionality between current density and intensity of applied electric field is an experimental result first formulated by George Ohm and named after him as **Ohm's law** for metallic conductors. Coefficient λ is called **specific conductivity** of a considered metal, from classical theory it follows that

$$\lambda = q \cdot n \cdot \mu = \frac{q^2 n \tau}{2m}$$

where τ is a mean time between collisions of electrons with ions of crystalline lattice (it is proportional to the mean free path of conducting electrons considered as gas particles), *m* is the mass of an electron.

Electric field with intensity *E* when moving one electron at a distance *dx* performs mechanical work equal to

$$dA_e = F \cdot dx = q \cdot E \cdot \bar{v} \cdot dt$$

The work per unit time (mechanical power) is

$$A_e = q \cdot \mu \cdot E^2$$

so the work being performed over all free electrons in a unit volume of a metal is

$$A = q \cdot \mu \cdot n \cdot E^2 = \lambda \cdot E^2$$

Phase state of a metal does not change if field intensity is not extremely high, so this work is completely converted into heat, and we can write that amount of heat released in a unit volume of a metal in conduction of electric current is equal to

$$Q = \lambda \cdot E^2 = \frac{j^2}{\lambda} \qquad (11)$$

Formula (11) is called the **Joule law**.

Ohm's law (10) can be rewritten in another way, using the concept of the **specific resistance** $\rho = 1/\lambda$:

$$j = \lambda \cdot E = \frac{E}{\rho}$$

Expressions (10) and (11) represent **differential forms** of appropriate laws. They state dependences between physical values in a point. Another form called **integral** one is also possible. It is widely used in electric engineering in problems, where cross-sectional area of metallic conductors is constant. Then Ohm's law for a metallic wire of length *l* and cross-sectional area *S* can be rewritten so:

$$I = j \cdot S = \frac{S}{\rho} \cdot E = \frac{S}{l \cdot \rho} \cdot E \cdot l = \frac{S}{l \cdot \rho} \cdot V$$

In the latter expression we used formula (6), *V* is the absolute value of electric potential difference between ends of the wire or **voltage**. Introducing value of electric resistance of a conductor $R = (\rho \cdot l)/S$, we obtain the Ohm law in the integral form:

$$I = \frac{V}{R} \qquad (12)$$

Voltage is measured in *Volts* (*V*), current intensity in Amperes (*A*), resistance in *Ohms* (Ω). The integral form of the Joule law can be derived too:

$$Q = U \cdot I = I^2 \cdot R \qquad (13)$$

6. Conservation of electric charge

The conservation of charge law states that difference between positive and negative electric charges in any closed volume could be changed only with moving some extra charge from the ambient space. This is the result of experimental observation that electric charges are always

generated in pairs of equal positive and negative particles. To formulate this law mathematically let us consider a very thin and short round cylinder with axis directed along the *x*-axis of a coordinate system (*Fig. 5*).

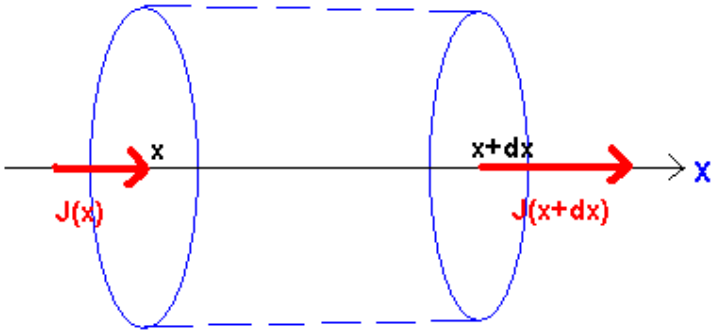


Fig. 5

Radius of the cylinder is *r*, its length is *dx*. Electric current flowing into the cylinder (suppose it can be only through the round sides) is equal to

$$I(x) = \pi r^2 j(x)$$

j(x) is the current density at some point *x*. Electric current leaving the cylinder at the position *x* + *dx* is

$$I(x+dx) = \pi r^2 j(x+dx)$$

Difference of this two values is the rate of total gain (or loss, depending on its sign) of charge in unit time

$$\frac{\partial Q}{\partial t} = \pi r^2 j(x+dx) - \pi r^2 j(x) = \pi r^2 \frac{\partial j}{\partial x} dx$$

If (for simplicity) to suppose that all charges have the same value, than the total charge is easily expressed through the carriers' concentration:

$$dQ = qn\pi r^2 dx$$

If to denote electric charge in a unit volume as $\rho = qn$, we obtain the differential form of the charge conservation law:

$$\frac{\partial \rho}{\partial t} = \frac{\partial j}{\partial x}$$

7. Electromotive force

We have mentioned previously that mechanic work performed on motion of an electric charge in a static (constant in time) field depends on values of electric potential in an initial and in finale points of motion. It can be said that this work does not depend on a trajectory of the motion. That leads to conclusion that no mechanic work is needed to move a charge along a closed path in electrostatic field, because potential difference is equal to zero in this case.

But we know that some heat released in conduction of electric current in metals. As it proved in thermodynamics process of heat release is irreversible, so energy spent for heat cannot be transform into energy of electric field in any reverse process. Then how to provide a constant current flow in electric circuits?

This problem is solved by use of different processes of non-electrostatic origin, called **electromotive forces (e.m.f.)**. Imagine, for example, a vessel divided by halves with porous semipermeable wall. If we pour some amount of hydrochloric acid into one half of the vessel, small ions of hydrogen would pass it much sooner than chlorine ions. Therefore after some time because of diffusion concentration of positive ions of hydrogen will be more in one half of the vessel, while distribution of Cl^- ions in vessel parts will be inverse. So one half of the vessel will be charged positively, opposite one – negatively. If we switch this vessel into electric circuit it would be able to separate electric charges to provide direct electric current in the circuit.

That is an example of electromotive force using diffusion (a chemical battery). There are some other physical processes, which are used as sources of electromotive forces, like photoelectric effect, electromagnetic induction, radioactivity and some others.

8. Kirchhoff rules for electric circuits

The first Kirchhoff rule for electric current in fact is a consequence of conservation of charge. Consider a node in an electric circuit, where some wires with electric current are connected together. If we denote values of income currents with sign “+” and values of outcome ones with “-“ (inverse designation is also valid), then algebraic sum of all currents passing through the node is equal to zero:

$$\sum_{i=1}^{n}\left(i_1 + i_2 + \ldots + i_n\right) = 0$$

Why? Because there is no generation or absorption of charge in any node of an electric circuit. The latter equation is a mathematical formulation of the first Kirchhoff rule.

The second Kirchhoff rule states that algebraic sum of all e.m.f. in every closed circuit is equal to algebraic sum of products of all resistances in that circuit by passing through those resistances currents.

An example. Let us consider a circuit drawn at *Fig. 6*.

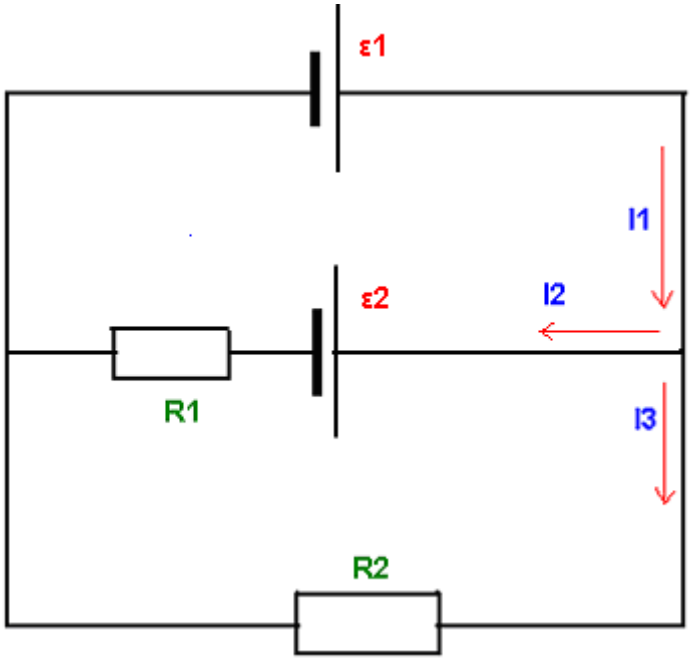


Fig. 6

According to the first Kirchhoff rule

$$I_1 = I_2 + I_3$$

According to the second one

$$\varepsilon_1 - \varepsilon_2 = I_2 R_2 \quad \text{and} \quad \varepsilon_1 = I_3 R_2$$

E.m.f. ε_1 and ε_2 are included with opposite signs because they are switched in opposite directions with respect to direction of the flowing current. So we obtained three algebraic equations for determination of three unknown currents (values of all e.m.f. and resistances are usually known). It is seen that one circuit could be chosen in different way, but the result, of course, would be the same.

9. Electric current in liquids and gases

Many liquids can conduct electric current, because their molecules can dissociate into positive and negative ions. These conducting liquids are called **electrolytes**. Electric current in electrolytes is transmitted with ions of both signs. Ohm's law is also valid in this case, but obeys a bit different formula, than in metals:

$$j = \left(q_+ \cdot n_+ \cdot \bar{v}_+ - q_- \cdot n_- \cdot \bar{v}_-\right) = \left(q_+ \cdot n_+ \cdot \mu_+ - q_- \cdot n_- \cdot \mu_-\right) \cdot E = \lambda \cdot E \quad (11)$$

Electron current in metals does not transmit substance. Electrons being gone as they reach the positive pole of a battery are substituted with other electrons from the negative pole. As Riecke had stated in his experiment (1901), current in metals is not accompanied by motion of ions of a crystalline lattice. Though electric current in electrolytes moves ions of solution, they settle at appropriate poles of an e.m.f. sorce. Michael Faraday, who studied this phenomenon experimentally, formulated the **law of electrolyses** named after him: *mass of a sediment at electrodes is proportional to the current intensity and to the time duration of the process*:

$$m = K \int I(t) \cdot dt \quad (12)$$

Coefficient of proportionality is

$$K = \frac{A}{F \cdot Z}$$

where A is the atomic mass of an ion, Z is its valence, $F \approx 9.6487 \cdot 10^4$ $C/mole$. In a particular case of a direct current equation (12) can be rewritten as

$$m = \frac{A \cdot Q}{F \cdot Z}$$

Usually dissociation of molecules of liquids into ions takes place without external electric field. Opposite situation takes place in gases. At normal conditions gas molecules are neutral, but external perturbations like heating or irradiation can release some electrons of outer shells so some amount of positive ions are created. Energy needed to ionize a molecule of a gas is called **energy of ionization** E_i**,** its value is of the order of 4 to 25 *eV* (*electron-Volt* is an off-system unit of energy, equal to energy in *Joule*

divided by value of charge of electron in *Coulomb*). Sometimes value of **ionization potential** is also used, it is equal

$$\varphi_i = \frac{E_i}{q}$$

Application of electric field leads to transition of electric current in such ionized gas. Current density in this case obeys the same formula (11), in which sign "-" denotes not negative ions like in liquids, but electrons. It was found experimentally that Ohm's law with its linear dependence between applied filed and resulting current flow density is valid in gases only at rather low electric field. Typical dependence of intensity of electric current flowing in gas on applied voltage is drawn at *Fig. 7*, it is often called volt-ampere characteristic or current-voltage characteristics of electric current in gases. (Usually such curves are measured for gas enclosed in some glass tube containing two metallic electrodes, to which voltage is applied. Such tube can be subjected to heating or illumination with x- or gamma-rays to form initial number of charge carriers.)

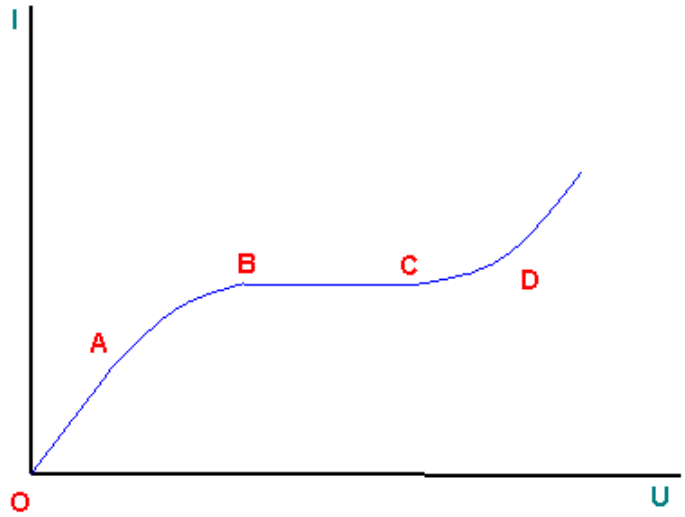


Fig. 7

Ohm's law is valid only at a short initial part *OA* of volt-ampere characteristics. Increase of electric filed accelerates electrons and ions to higher values of their velocities. Some of them reach kinetic energy exceeding energy of ionization and colliding neutral gas molecules, generate new electron-ion pairs, which continue this process multiplying concentration of charges. As concentration of ions and electrons increases collisions between them take place more often and growing process of recombination (reverse combination of molecules from charged particles)

diminishes rate of increase current with increase voltage at the piece *AB* of *Fig. 7*. At some voltage all molecules of gas become ionized and current does not grow with increase of electric filed (piece *BC* of volt-ampere characteristics). Further increase of voltage applied to a gas tube leads to the steep increase if curve (after the point *D*). It had been found that as value of current at the point *D*, external field can be removed without influence on value of electric current. This is the **self-sustained gas discharge**, which is maintained with generation of very large amounts of charges by continuous bombardment of molecules with electron and ions.

Electric current in gases depending on values of gas pressure and of applied voltages can be observed in quite different forms. **Glow discharge** used in daylight lamps takes place at very low pressure values (of order of some *mm of mercury*) and voltages of some hundreds *V*. Radiated energy is created by transitions of electrons from higher to lower energy levels in processes of bombardment of molecules and ions by accelerated electrons. Recombination of ions and electrons is another one source of luminous energy in glow discharge.

Spark discharge, which could be seen as lightings during thunderstorms, takes place at normal atmospheric pressure as very high voltage (of the order of dozens of *kV*) is formed between clouds, for example. Lightings are usually seen as of zigzag lines (spark channels) and are accompanied by loud sound. It takes place because as very strong current passes between two points in a gas electric conductivity of that part of air abruptly diminishes that leads to rapid increase of current, rapid heating of appropriate volume of the gas, therefore to rapid growth of pressure with generation of an acoustic shock.

Corona discharge is often observed at atmospheric pressure in very non-uniform electric field around thin electric wires. We know from formula (6) that intensity of electric field is proportional to rate of potential in space. Metallic tips have small radii, near which gradient of electric potential can reach very large value, resulting strong electric field ionizes surrounding air. Recombination of ion-electron pairs and transitions of electrons from excited to ground energy levels radiate visible light. To prevent loss of electric energy for corona discharge electric transmission line are usually made from thick wire.

Spark discharge between electrodes placed at very small distance (of the order of some millimeters) is called **voltaic arc** or **arc discharge**.

10. Plasma

Plasma is the approximately electrically neutral gas of ions and electrons. It is selected as a special phase state of substance because many

processes in plasma take place differently with respect to a conducting medium containing opposite electric charge, but as in a medium with collective behavior of electrons and ions. For example, if concentration of particles in plasma is rather low (collisions are sufficiently rare), it is possible that average kinetic energy of electrons has one value, kinetic energy of ions has another value, while kinetic energy of unionized atoms has some other value different from both above mentioned. That means that this kind of plasma has 3 different character temperatures: electrons' temperature, ions' temperature and atoms' temperature simultaneously. A special branch of physics is devoted to processes observed in plasma. Except for plasma channels in glow or spark discharges, which have been mentioned in the present Lecture, all shining stars consist of plasma, so we can say that plasma is the prevailing state of matter in our Universe.

LECTURE 6. MAGNETIC FIELD. ALTERNATING CURRENT.

1. Moving electric charges and magnetic field

Antique historians left records about ancient Phoenicians who used pieces of natural magnets for orientation in their navigation in Mediterranean sea. Attraction of iron objects with permanent magnets also was observed more than 2000 years ago. But only in the beginning of XIX century it had become evident that there is some relation between magnetic properties of substances and motion of electric charges.

A Danish physicist Hans Christian Oersted (Ørsted in Danish) placed a magnetic needle, being fasten at a vertical axis, near a thin metallic wire (*Fig. 1*). This demonstration first took place in 1820.

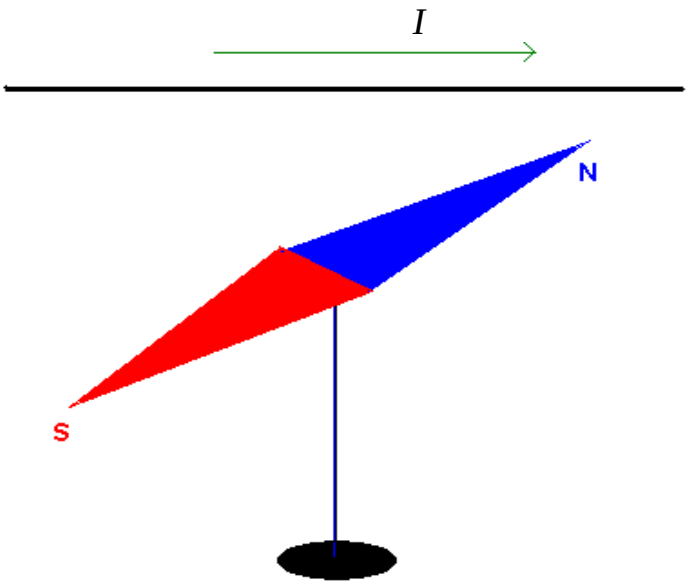


Fig. 1

Switching electric current he caused the needle to rotate perpendicularly to the wire, if direction of the current was changed to opposite, the magnetic needle also changed its direction to the opposite one. This and some other experimental observation helped to formulate two basic laws of interaction currents and magnetic fields. French investigators Biot and Savart have found that an electric current of intensity I flowing through a wire of length l generates magnetic field of **inductance** equal to

$$\vec{B} = \frac{I \cdot \mu_0}{r^3} \cdot \left(\vec{l} \times \vec{r}\right) \qquad (1)$$

where vector $\vec{l}$ is directed along the wire, vector $\vec{r}$ passes from the wire to a point, in which value of the magnetic field is determined, $\mu_0 = 4\pi \cdot 10^{-7}\ H/m$ is the so-called magnetic permeability of a vacuum. Vector cross-product in formula (1) means that maximal value of magnetic induction takes place when vectors $\vec{l}$ and $\vec{r}$ are perpendicular to each other (vector $\vec{B}$ is perpendicular to both them). Correspondingly, if vectors $\vec{l}$ and $\vec{r}$ are parallel to each other, then vector $\vec{B}$ is equal to zero. Vector $\vec{B}$ in fact is intensity of the magnetic field, but it is called inductance only because of tradition. Magnetic induction in the SI system is measured in *Tesla* (abbreviation is *T* or *Tl*). $1T = 1N/(A \cdot m)$.

In the same first decades of XIX century Andre-Mary Ampère discovered that mechanical force acting at electric current of intensity *I* flowing through the wire of length *l* obeys the next formula:

$$\vec{F} = I \cdot (\vec{l} \times \vec{B}) \qquad (2)$$

A cross-product of vectors on the right-hand side of (2) points out that value of the force essentially depends on mutual orientation of the current and magnetic field. As electric current is a motion of charges, formulas (1) and (2) can be rewritten for a single charge moving with velocity *v*:

$$\vec{B} = \frac{q \cdot \mu_0}{r^3} \cdot (\vec{v} \times \vec{r}) \qquad (1')$$

$$\vec{F} = q \cdot (\vec{v} \times \vec{B}) \qquad (2')$$

If to add the Coulomb law to the right-hand part of (2') we could get the expression for **Lorentz force** acting at a charge in combined electric and magnetic field:

$$\vec{F} = q \cdot (\vec{E} + \vec{v} \times \vec{B}) \qquad (3)$$

It is seen from (2') if we consider a charge in a frame, in which it is immovable (we move parallel to it with the same speed), then magnetic (Ampère) force vanishes, but electrostatic Coulomb force remains.

You should also know how expressions (1') and (2') look in Gaussian system of units (more preferable in electricity and magnetism):

$$\vec{B} = \frac{q}{c \cdot r^3} \cdot (\vec{v} \times \vec{r}) \qquad (1'')$$

$$\vec{F} = \frac{q}{c} \cdot \left(\vec{v} \times \vec{B} \right) \qquad (2'')$$

Magnetic permeability of vacuum is not included in formulas (1'') and (2'') (as well as electric permeability does in Gaussian system in electrostatics), but speed of a charge is changed everywhere with its ratio by speed of light in vacuum c.

Magnetic field also can be drawn graphically as a picture of field curves, which start from northern poles of magnets and end at south poles (*Fig. 2*). But, contrary to the electric field, magnetic field lines are always closed, because there no magnetic separated positive and negative magnetic charges. Strictly speaking, according to the contemporary theory we believe that at the first short time interval after the Big Bang magnetic charges existed, but they annihilate (annihilation is a junction of two particles of opposite sign with theirs' vanishing and releasing some amount of energy) very quickly. Their concentration nowadays is very small though no magnetic monopoles are detected experimentally up to now.

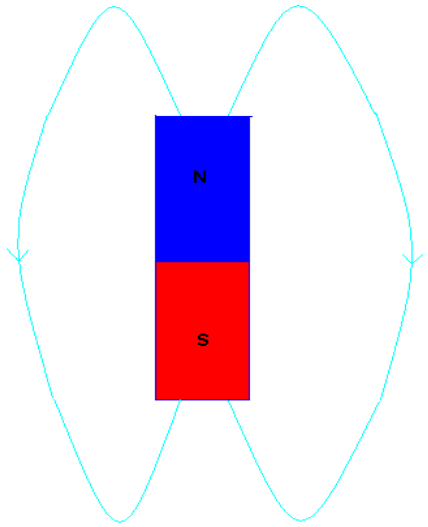


Fig. 2

2. Current loop in a magnetic field

Let us consider a rectangular closed loop *abcd*, rotating around the axis, which is perpendicular to vector of induction of magnetic filed (*Fig.3*). Direct electric current of intensity *I* is flowing in the loop. According to the Ampère's law (2) mechanical forces act at all sides of the loop. Forces applied to the sides *ba* and *dc* are oppositetely directed along the axis of rotation, they are equal at any moment of time. Forces applied to the sides

cb and *ad* are also compensate each other, though they produce a mechanical torque, which rotates the loop around the axis.

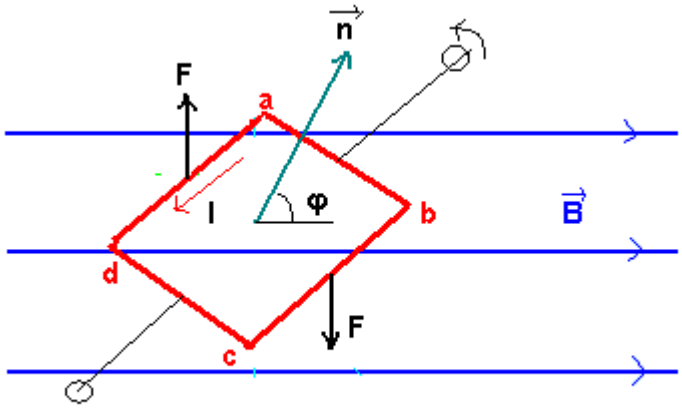


Fig. 3

This torque is equal to

$$M = 2F \cdot \frac{h}{2} \cdot \cos\varphi$$

where φ is the angle between the normal to the plane of the loop and the direction of the magnetic induction, *h* is the length of the side *cd* or *ab* of the rectangular. The normal to the loop is to be drawn so that from its tip current flow in the loop would be seen counterclockwise. The latter formula can be rewritten using the Ampère's law:

$$M = I \cdot B \cdot S \cdot \cos\varphi \qquad (4)$$

where *S* is the area of the loop. The considered example is very important for explanation of many magnetic phenomena, therefore let us separate values describing field and values characterizing the loop in (4). For this purpose we introduce a value of a **magnetic moment** of the loop $\vec{P}$ (direction of a vector of magnetic moment of a loop always coincides with the direction of a loop's normal $\vec{n}$):

$$\vec{P} = I \cdot S \cdot \vec{n} \qquad (5)$$

Using (5) we can rewrite (4) in the next way:

$$M = \vec{P} \cdot \vec{B} = |\vec{P}| \cdot |\vec{B}| \cdot \cos\varphi \qquad (6)$$

here φ is the angle between vectors $\vec{P}$ and $\vec{B}$.

Another one useful value, which can be introduced for a current loop in a magnetic field is the **magnetic flux** through the loop, which is equal to

$$\Phi = \left(\vec{B} \cdot \vec{n}\right) \cdot S \tag{7}$$

so that the considered torque can be expressed though the magnetic flux too:

$$M = I \cdot \Phi \tag{8}$$

In the SI system magnetic flux is measured in *Weber* (*Wb*). It is seen that $1Wb = 1T \cdot m^2$.

3. Electromagnetic induction

Consider an U-shaped metallic frame and a straight metallic bar *ab*, which can slide without friction along parallel sides of the frame (*Fig. 4*). Suppose the plane of this circuit is perpendicular to the direction of a vector of magnetic induction $\vec{B}$. If battery, generating e.m.f. ε, is switched into the frame, then electric current with intensity I flows in this closed loop. Suppose we move the bar with velocity $\vec{v}$ parallel to its initial position. According to the Ampère's law a force $F = q \cdot v \cdot B$, directed from *b* to *a*, acts at every positive charge in the bar (here l is a distance between *a* and *b*). As current carriers in metals are electrons, a force acting at every electron is directed oppositely *a* to *b*.

In other words, one more e.m.f. is appeared in the loop, absolute value of which is equal to

$$\varepsilon = E \cdot l = \frac{F \cdot l}{q} = v \cdot l \cdot B \tag{9}$$

Comparing the latter formula with equation (7), we can easily see that the right-hand side of (9) is the rate of change of the magnetic flux, penetrating the loop.

It was Michael Faraday who first stated in 1831 that temporal change of a magnetic flux through some loop generates electromotive force in the loop, which tends to diminish influence of that change or, in other words, this e.m.f. tends to maintain magnetic flux constant.

$$\varepsilon = -\frac{d\Phi}{dt} \qquad (10)$$

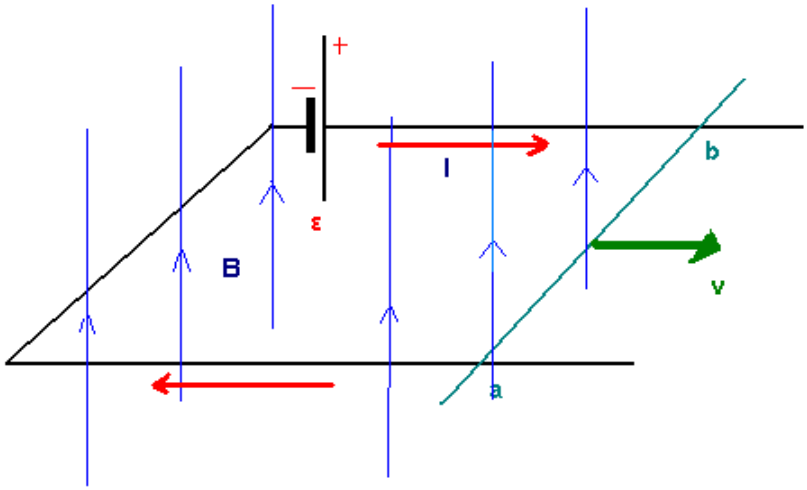


Fig. 4

Formula (10), neglecting evident simplicity of its derivation, is very important, it is one Maxwell's laws of elecrodynamics, which state fundamentals of our knowledge on electricity and magnetism.

4. Magnetic field in substances

Everybody knows that atoms consist of tiny positively charged nuclei and negative electrons. In a semiclassical model we consider electrons as charged dots rotating around nuclei. But rotating charge is electric current with density equal to charge value multiplied by frequency of rotation:

$$i = q \cdot f = q \cdot \frac{v}{2\pi r}$$

where r is the radius of electron's orbit, v is the linear speed of its orbital motion.

Magnetic moment of a rotating electron can be found from Eq. (5)

$$\vec{P}_o = q \cdot f \cdot S \cdot \vec{n} = q \cdot f \cdot \pi \cdot r^2 \cdot \vec{n} = \frac{1}{2} q \cdot v \cdot r \cdot \vec{n}$$

Besides that, as planets can rotate around their own axis together with rotation around their stars, we can imagine rotation of an electron

around the axis passing through its center. The former rotation is called **spin**. This is only a classical analogy using for explanation, real electron is more similar to a cloud distributed around the nucleous and the spin is one of properties of this cloud. In any case, spin of an electron also makes contribution into its magnetic moment. It is calculated in quantum mechanics that its value is

$$\left|\vec{P}_s\right| = \frac{q \cdot h}{4\pi \cdot m_e}$$

h is the Planck's constant, m_e is the mass of an electron. It is interesting, that the latter expression contains only constants and it is called **Bohr magneton** (see also Lecture 14)

$$\mu_B = (q \cdot h)/(4\pi \cdot m_e) \approx 9.27 \cdot 10^{-24}\ J/T \qquad (11)$$

so that

$$\left|\vec{P}_s\right| = \mu_B$$

Total magnetic moment of an atom is the vector sum of orbital and spin magnetic moments of all its electrons and of spin moments of all its protons and neutrons (which also have spins). However magnetic moments of protons and neutrons are usually much less then those of electrons and could be neglected, so total magnetic moment of a unit volume of a substance is

$$\vec{I} = \frac{\sum \vec{P}_o + \sum \vec{P}_s}{V}$$

V is volume. Vector $\vec{I}$ is called **magnetization** of a substance, measured in A/m.

If substance has been placed in the external magnetic field of induction $\vec{B}_0$, then total induction inside it is equal to

$$\vec{B} = \vec{B}_0 + \vec{I}$$

It was found experimentally that in many cases magnetization of a substance is proportional to the value of applied field $\vec{I} = \chi \vec{B}_0$ (χ is the

magnetic susceptibility of a substance). So induction of magnetic field in the substance is

$$\vec{B} = \vec{B}_0 + \vec{I} = \vec{B}_0 + \chi\vec{B}_0 = \mu\vec{B}_0 \qquad (12)$$

where μ is the **magnetic permeability** of a substance. To characterize magnetic field inside substances a vector of **intensity of magnetic field** $\vec{H}$ is introduced along with magnetic induction:

$$\vec{B} = \mu_0\mu\vec{H} \qquad (13)$$

It should be mentioned that we consider hear the simplest case of materials with rather weak magnetic properties, in which magnetic susceptibility and permeability do not depend of a value of applied field. In ferromagnets, which will be considered soon this is incorrect and

$$\chi = \chi(\vec{B}_0) \quad \text{and} \quad \mu = \mu(\vec{B}_0)$$

wherefore directions of the external and internal magnetic fields in a ferromagnet do not coinside in general case.

Subject to value of magnetic susceptibility (or permeability) all substances are devided into some kinds.

Diamagnets are substances, in which χ is negative and of the order of 10^{-6}–10^{-4}, Water, proteins, gold, silver and some other metals are diamagnets.

In fact, all substances have diamagnetic properties. This the consequence of electromagnetic induction, impeding increase of the external field. But atoms of some substances have their own magnetic moments without external field. At normal temperatures those magnetic are directed randomly in different directions. Application of an external magnetic forces them to align in the field's direction. Those are **paramagnets** ($\chi \sim +(10^{-6}$–$10^{-4})$). Air, oxygen, nitrogen, alkali metals are paramagnets.

Ferromagnets without the external field contain small, but macroscopic (with size of the order to some millimeters) parts with aligned in one direction magnetic moments. These parts of a ferromagnet with a constant directions of magnetic moment are called **domains**. External magnetic field arrange them, so that value of the field within a ferromagnet can be some hundreds or even thousands times more than the applied one. Eight metals (*Fe, Ni, Co, Gd, Dy, Ho, Tb, Er*) and their numerous alloys are belong to ferromagnetics. Typical curves of dependence of inductance of

the magnetic field $\vec{B}$ in a ferromagnet on intensity of external field $\vec{H}$ are drawn at *Fig. 5*

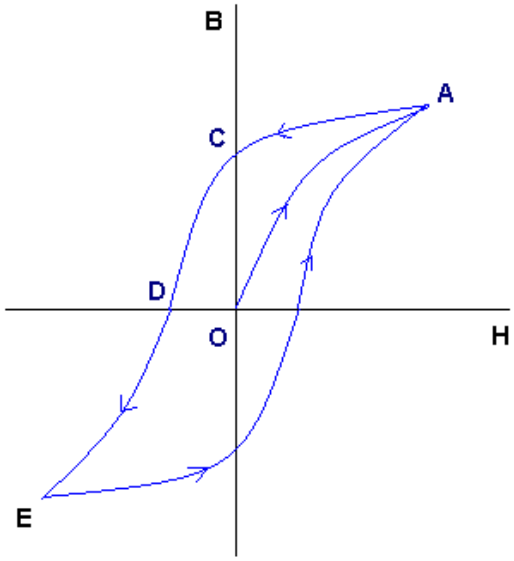


Fig. 5

If a ferromagnet was not magnetized initially then inductance grows with the growth of intensity along the curve *OA*. It is seen that this growth of B becomes slower and soon after point *A* increase of intensity does not lead to increase of inductance because of saturation – all domains are oriented in the direction of the external field. Diminishing of intensity to the zero value does not eliminate magnetic field inductance in a ferromagnet: the point *C* is the residual value (or **remanence**) of internal inductance. Further growth of applied intensity in opposite to initial direction from the point *C* leads to final elimination of the magnetic inductance at some value of intensity denoted with the point *D*. This value of intensity is called the **coercive force (**or **coercivity)**. Due to existence of residual intensity of magnetic field in ferromagnets, they are used as permanent magnets. Such dependence of one physical value on another, when the prehistory of should be taken into account, is called **hysteresis**.

5. Alternating electric current

According to definition, **alternating current** is a directional motion of electric charges intensity of which is a harmonic function of time ($\sim \sin \omega t$ or $\sim \cos \omega t$). Any electric current is a result of application of voltage (potential difference) between different points in a conducting medium. So we can suppose that voltage

$$V(t) = V_0 \cos(\omega t + \varphi_0)$$

where similarly to the theory of mechanical oscillations ω is called the cyclic frequency of voltage, φ_0 is initial phase. $V(t)$ is an instant value of voltage, V_0 is its amplitude (maximal value).

If alternating voltage is applied to an active (Ohmic) resistance then like in the case of direct current

$$I(t) = \frac{V(t)}{R} = I_0 \cos(\omega t + \varphi_0)$$

Alternating current passing through an Ohmic resistance is always in phase with applied voltage, it is expressed with the same value of initial phase in the above mentioned formulas.

If some other loads are switched in a circuit with alternating voltage the Ohm's law remains correct, namely

$$I(t) = \frac{V(t)}{Z(\omega)} \qquad (14)$$

though in general case inverse coefficient of proportionality between voltage and current is called **electric impedance** and its value can determine of frequency.

Electric impedance of a circuit depends on frequency if the circuit contains some condensers (capacitors) or coils. First condensers were designed in XVII century to store electric charge. The simplest condenser is a pair of two parallel flat metallic electrodes separated with a thin layer of gaseous or solid dielectric. Ability to accumulate electric charge is determined by a physical value called capacity. Capacity of a flat condenser is proportional to area of its plates *S* and inversely proportional to distance between them *d*:

$$C = \frac{\varepsilon\varepsilon_0 S}{d}$$

where ε is the dielectric permittivity of a dielectric between plates. In the SI system capacity is measured in *farad* (*F*). If the condenser is charged, then voltage between plates is proportional to the stored charge for a condenser of any shape:

$$V = Q/C \qquad (15)$$

Direct electric current cannot pass through a condenser, because there are no current carriers in a dielectric between its plates. Nevertheless, if a capacitor is turned on to a battery, free electrons from one condenser's plate immediately move to he positive pole of the battery, leaving some amount uncompensated positive ions in a crystalline lattice of the plate. This positive charge attracts some electrons from the negative pole. This process continues until voltage between every plate and the appropriate pole of the battery becomes zero, really it takes some micro- or at most milliseconds. After it has been finished the capasitor is charged.

If a condenser is turned on into the circuit with a source of alternating voltage, then those plates are recharded every half of period of oscillations of voltage. In this respect we usually say that alternating current passes through capacitors, though, of cause, no charge carriers cross the the gap between plates, because there are no carriers.

Using the latter formula for voltage at a condenser and Ohm's law for a load in an alternating current mains, we have

$$I(t) = \frac{V(t)}{Z(\omega)} = \frac{dQ}{dt} = CV_0 \frac{d\left(\cos(\omega t + \varphi_0)\right)}{dt} = -C\omega V_0 \sin(\omega t + \varphi_0) = C\omega V(t) \tag{16}$$

Comparing expressions on the right-hand side of (14) and (15) we can find that electric impedance of a condenser is equal to

$$Z_C = 1/\omega C$$

Besides that, it seen that voltage at a condenser is not in phase with passing through it electric current, because minimal and maximal values of *sin* and *cos* functions of the same argument take place in different moments of time. It could be easily understood, if we recall recent description of a process of charging of a capacitor. Indeed, as it is uncharged initially voltage between plates is zero (look at (15)), but voltage between every plate and the appropriate pole of a battery is maximal, therefore current intensity is maximal at this moment of time. As the capacitor becomes charged, there is maximal voltage between its plates, but intensity of current is zero.

Another one widely used kind of load in an alternating current mains is a solenoid (a coil) consisting on some amount of rings of a wound wire in the simplest case (*Fig. 6*)

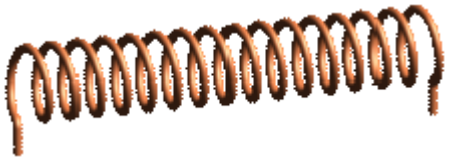


Fig. 6

Before consideration electric current in a coil, let us recall a closed loop with current. We know that magnetic flux through the loop is equal to

$$\Phi = B \cdot S \cdot \cos\varphi$$

where φ is the angle between vectors of magnetic inductance $\vec{B}$ and normal to the plane of the loop. Using Biot-Savart law it is possible to calculate that magnetic inductance generated in the loop by its own current I is proportional to current intensity (coefficient of proportionality depends on a point, in which magnetic inductance is measured):

$$B = k \cdot I$$

Therefore magnetic flux passing through the plane of the loop with current is also proportional to current intensity:

$$\Phi = L \cdot I$$

Coefficient L is called **inductance**, it is measured in *henry* (*H*) in the SI system.

If intensity of the current in the loop is changing, then e.m.f. of electromagnetic induction is being generated:

$$\varepsilon = -\frac{d\Phi}{dt} = -L\frac{dI}{dt} \qquad (17)$$

Sign minus means that this e.m.f tends to prevent change of magnetic flux or in the case under consideration to prevent change of current's strength.

The same phenomenon would be considered in a long coil (solenoid), its inductance is equal to

$$L = \frac{\mu\mu_0 SN^2}{l}$$

where μ is the magnetic permeability of a medium between turns of the coil, S is the area of the turn, l is the length of a solenoid, N is the number of its round turns.

Thus voltage drop on a coil is equal to

$$V(t) = -L\frac{dI}{dt}$$

Ohm's law for a coil in an alternating current mains looks like

$$I(t) = \frac{V(t)}{Z(\omega)} = -\frac{V_0}{L}\int \cos(\omega t + \varphi_0)dt = -\frac{V_0}{\omega L} \qquad (18)$$

So we can conclude that $Z_L = \omega L$.

If we consider all three loads (resistance, condenser and coil) switched in the same circuit in series, then total voltage drop, which is equal to applied alternating voltage (initial phase shift is chosen zero for simplicity):

$$V(t) = V_0 \cos\omega t$$

would be equal to sum of three drops:

$$V_0 \cos\omega t = I \cdot R + \frac{Q}{C} + L\frac{dI}{dt}$$

As electric current strength is derivative of charge on time both parts of this equation can be differentiated on time, to obtain:

$$-\omega V_0 \sin\omega t = L \cdot \frac{d^2 I}{dt^2} + \frac{dI}{dt} \cdot R + \frac{I}{C} \qquad (19)$$

From mathematical point of view this the same equation as in formula (6) of Lecture 2, where we considered forced mechanical oscillations. So it has a similar solution

$$I = I_0 \cos(\omega t + \varphi) \qquad (20)$$

Using Ohm's law it can be found that impedance of such RLC circuit is equal to

$$Z = \sqrt{R^2 + \left(\omega L - 1/\omega C\right)^2} \qquad (21)$$

while the tangent of a phase shift between voltage and current in (20) is

$$\tan\varphi = \frac{\omega L - 1/\omega C}{R} \qquad (22)$$

It is seen from (21) that impedance of such circuit is minimal if expression in brackets is zero. It means that at the frequency

$$f = \frac{1}{2\pi\sqrt{LC}} \qquad (23)$$

electric current in an RLC circuit achieves its maximal value for the same applied voltage. This phenomenon is called **resonance** like in theory of mechanical oscillations (Lection 2).

Nowadays absolute majority of consumers of electric energy are supplied with alternating, but not direct voltage because transportation of the alternating electric current at long distances leads to much lower power loss than transportation of the direct current. Besides that, change of voltage from a transmission line to recipients is much easier in the case of alternating voltage. Alternating voltage is usually generated at electric power plants in electric generated principally designed as a conducting loops rotated with constant angular speed in magnetic field (like at *Fig. 3*). Alternating voltage is a result of electromagnetic induction in this case.

6. Joule law in an alternating current circuit

To determine amount of heat released in a circuit as alternating current passes we calculate product of the current strength and applied voltage during a period of oscillations and divide this value by the period:

$$w = \frac{1}{T}\int_0^T V_0 \cos(\omega t)\cdot I_0 \cos(\omega t + \varphi)\cdot dt = \frac{V_0 I_0}{2} \qquad (24)$$

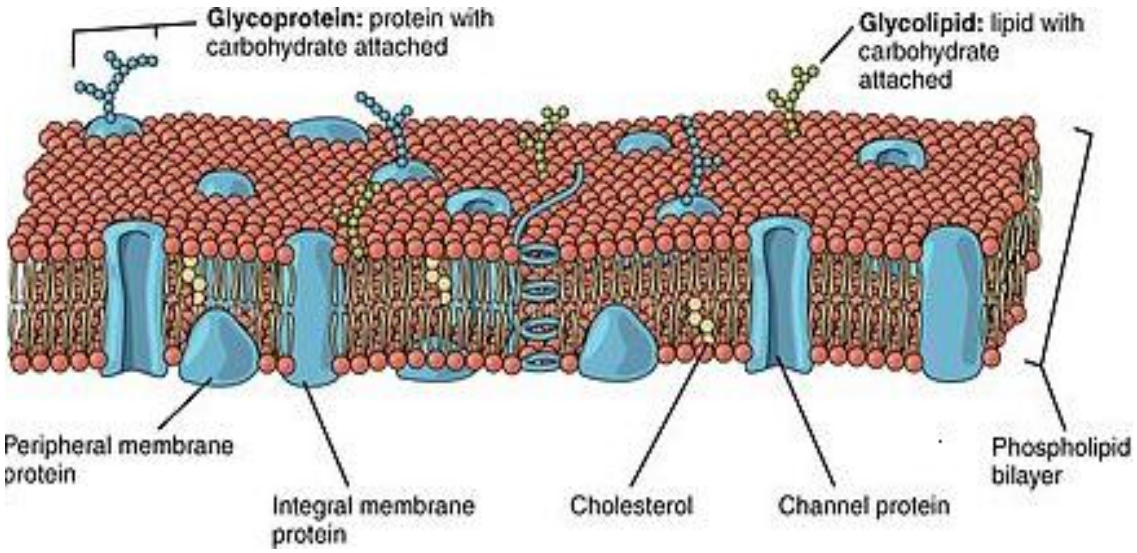
To fashion this formula into a more habitual form, notions of **effective voltage** and **effective current** strength are used:

$$V_{eff} = V_0/\sqrt{2};\ \ I_{eff} = I_0/\sqrt{2} \qquad \text{giving} \quad w = V_{eff}\cdot I_{eff} \qquad (25)$$

LECTURE 7. PHYSICS OF BIOLOGICAL MEMBRANES

1. Functions and structure of biological membranes

This lecture is devoted to biological membranes surrownding cells in living tissues. All those membranes consist of two layers of phospholipid molecules with some molecules (usually of proteins) embedded between layers, the drawing at *Fig. 1* is downloaded from the Wikipedia as an illustration.



I

Fig. 1

A typical fragment of the double layer of **phospholipid molecules** is schematically drawn at *Fig. 2*. Every molecule contains as small dipole head (which is able to attract, for example, water because of dipole-dipole electrostatic interaction) and a pair of long tails of carbon atoms (from 14 to 24) with connected to them atoms of hydrogen. These tales repulse water dipoles, but using covalent bonds they attract each other laterally and between layers.

Fig. 2

Molecules of proteins in membranes are of two kinds. **Peripheral proteins** are penetrated into one phospholipid layer and can be easily removed. **Integral proteins** pass through the whole membrane, they play important role in transmission of substances inside a cell or outside it. There also can be cholesterol molecules drawn with yellow color at Fig. 1, they are character mostly to membranes surrounding hemoglobine cells, and molecules of glycoproteins and glycolipids (polymerized sugar attracted to a protein or a phospholipid molecule respectively). Sugars are usually connected only at one side of a membrane.

Membranes form cells' structure, protect them from external influence and ensure matter exchange with surrounding medium and supply with building material during cell division.

2. Some physical properties of membranes

Biological membranes are very thin, their usual thickness is of the order from some dozens of angstroms to a dozen of nanometers.

Biomembranes are lighter than water ($\rho \approx 800\ kg/m^3$). Therefore monolayer membranes used for experiment investigation can be obtained by collection of a thin phospholipid film from the water surface.

Electric conductivity of membranes is very low. Because of their very small thickness, they form very good capacitors. Dielectric permittivity of in the heads' part is of the order of $\varepsilon \sim 10$, while in the tails' part $\varepsilon \approx 2$.

Phospholipid as well as protein molecules can change their position in an appropriate layer (lateral diffusion), a phospholipid can move around the whole cell in approximately 1 second. Long phospholipid tails can rotate around their heads. Such fluidic behavior makes membranes similar to liquids in their properties, though arrangement of lipid molecules demonstrates long-range order. Therefore at normal conditions biological membranes are considered as **liquid crystals**. At lowering of temperature long lipid tails are erected, thickness of a membrane increases as well as a mean distance between molecules, a membrane turns into the **gel** state. In this state low-molecular substances can leave a cell through widen channels in the membrane. It can be important to prevent cells from destruction in the process of water cooling.

3. Simple diffusion through membranes. Fick equation for membranes.

Small-size molecules like water, oxygen or carbon dioxide move through biological membranes using simple diffusion. The Fick equation

was considered in the Lecture 3. We can recall that flux density of a substance (ions, molecules) with concentration *c* in the direction *x* is

$$j = -D\frac{dc}{dx}$$

For diffusion through membranes this equation is changed. Because of membranes' very small thickness ratio of differentials on the right-hand side of the latter equation can be substituted for the gradient:

$$j = -D\frac{c_{mi} - c_{mo}}{l}$$

where c_{mi} is the concentration in a membrane near the side adjacent to the cell, c_{mo} is the concentration in a membrane near the opposite side, l is the thickness of the membrane.

Concentrations in a cell or in a surrounding meadium could be measured more easily than those within a membrane. Therefore usually the next experimental fact is used: at the interface between two media concentration of a diffusing substance changes stepwise, and a value of **distribution coefficient** is introduced:

$$K = \frac{c_{mi}}{c_i} = \frac{c_{mo}}{c_o}$$

where the subscript *m* indicates values of concentration in a membrane (*Fig. 3* below). Thus we obtain the Fick equation for diffusion through membranes:

$$j = -DK\frac{c_i - c_o}{l} \qquad (1)$$

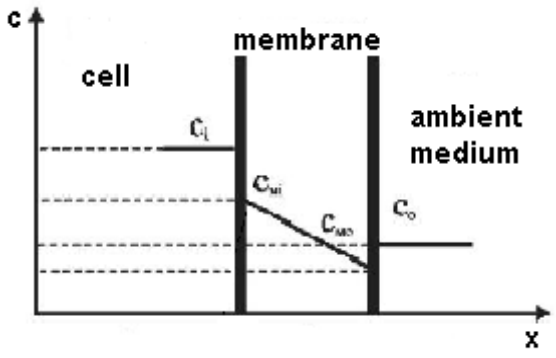


Fig. 3

4. Nernst-Planck equation for diffusion of charged particles through biological membranes

Some substances, like potassium, sodium, chloride, calcium cross biological membranes as ions. Therefore electric field in a membrane influence their transportation. A force acting at one ion carrying a charge ***Q*** in electrostatic field of intensity *E* is

$$f = Q \cdot E$$

If valence of the ion is Z then $Q = q \cdot Z$. Using the expression of field intensity the force can be rewritten as

$$f = -q \cdot Z \cdot \frac{\partial \varphi}{\partial x}$$

Friction force also acts at the ion, so we can assume that it moves with some average speed, which as we know from the previous lecture is proportional to electrostatic force (coefficient of proportionally below is the mobility of the ion μ):

$$v = \mu \cdot f = -q \cdot Z \cdot \mu \cdot \frac{\partial \varphi}{\partial x}$$

Total flux of a mole of ions is proportional to that speed, N_A (Avogadro number), as well as the concentration of ions *c*:

$$\Phi = c \cdot v \cdot S = -q \cdot Z \cdot \mu \cdot c \cdot N_A \frac{\partial \varphi}{\partial x} \cdot S = -Z \cdot \mu \cdot c \cdot F \cdot \frac{\partial \varphi}{\partial x} \cdot S$$

where $F = q \cdot N_A$ is the Faraday number, which is familiar to us from the Lecture 6. Flux density, according to definition, is equal to the latter value per unit area:

$$j = -Z \cdot \mu \cdot c \cdot F \cdot \frac{\partial \varphi}{\partial x}$$

In general transportation of ions consists of simple diffusion and drift in electrostatic field:

$$j = -D\frac{\partial c}{\partial x} - Z \cdot \mu \cdot c \cdot F \cdot \frac{\partial \varphi}{\partial x} \qquad (2)$$

Applying the Einstein's relation for coefficient of diffusion $D = \mu RT$, equation (2) can be rewritten as

$$j = -D \cdot \left(\frac{\partial c}{\partial x} + \frac{Z \cdot c \cdot F}{RT} \cdot \frac{\partial \varphi}{\partial x} \right) \qquad (3)$$

Formulas (2) and (3) are different variants of the **Nernst-Planck formula**.

5. Resting potential and Nernst formula

Voltage between the cytoplasm of a cell and ambient medium in absence of external excitation is called a **resting potential**. Since the beginning of the XX century resting potential was considered to form by potassium ions.

Using the Nernst-Planck formula (3) an expression for the resting potential can be readily obtained. In the state of rest there is no ions' flux through a membrane:

$$j = 0 \Rightarrow \frac{\partial c}{\partial x} + \frac{Z \cdot c \cdot F}{RT} \cdot \frac{\partial \varphi}{\partial x} = 0$$

So we have one differential equation with two independent variables. To simplify it, let us suppose that electric potential is zero at one side of the membrane and is equal to some resting value φ_r at the opposite side. Then we have an elementary ODE:

$$\frac{\partial c}{\partial x} + \frac{Z \cdot c \cdot F}{RT} \cdot \frac{\varphi_r}{l} = 0$$

It can be easily solved, so that concentration is equal

$$c = c_0 \cdot \exp\left(-\frac{ZF}{RT} \cdot \frac{\varphi_r}{l} \cdot x \right)$$

as at one side of the membrane ($x = l$) electric potential is φ_r, while at another side ($x = 0$) it is equal to zero, then from the latter equation follows

$$\varphi_r = -\frac{RT}{ZF} \cdot \ln \frac{\left[K^+\right]_{in}}{\left[K^+\right]_{out}} \qquad (4)$$

This is the **Nernst formula** for the resting potential. However, it was found that formula (4) gives overestimated value of a resting potential. Goldman, Hodgkin and Katz saved the situation, when proposed that concentrations of sodium and chloride ions also should be included in calculation. The next formula **named after Goldman, Hodgkin and Katz** is in much better agreement with experimental data:

$$\varphi_r = -\frac{RT}{F} \cdot \ln \left\{ \frac{P_K \cdot \left[K^+\right]_{in} + P_{Na} \cdot \left[Na^+\right]_{in} + P_{Cl} \cdot \left[Cl^+\right]_{out}}{P_K \cdot \left[K^+\right]_{out} + P_{Na} \cdot \left[Na^+\right]_{out} + P_{Cl} \cdot \left[Cl^+\right]_{in}} \right\} \qquad (5)$$

where every coefficient $P_m = (D_m \cdot K_m)/l$ is called the **penetrability factor** for the appropriate kind of ions.

6. Facilitated diffusion

Not very small molecules (amino acids, sugars) cannot penetrate membranes due to simple diffusion between phospholipid molecules. Simple non-mediated diffusion is also impossible for ions (K^+ and Na^+) because polar heads of lipid molecules repulse them. So-called **facilitated diffusion** through channels or with the help of carriers is used in such cases.

Channels for diffusion are formed from 4, 5 or 6 integral proteins arranged parallel to each other, so that a round fence of proteins with a small tunnel between them is made. This tunnel is filled with water and can be opened or closed after reception of a signal (change of membrane potential, for example). Ions when being transported through such channel can be surrounded with some water molecules, which leave "their" ion after transportation. Diffusion through channels is rather fast, about 10^8 ions per second can be transmitted in this way.

Another one mechanism of the facilitated diffusion is usage of **carriers** – protein molecules, which catch one appropriate ion or molecule and move them to the opposite side of a membrane. Maximal rate of this mechanism is lower than of channels – 10^4 particles per second.

Both mechanisms of facilitated diffusion differ from simple one in two main features: they are **selective** (a channel or a carrier able to transport one sort of ions never moves another one sort) and **saturable** (*Fig. 4*, the

straight line shows simple diffusion, the curve – facilitated one). Saturation can be reached when all channels or carriers of a needed kind are busy, so increase of ions' concentration does not accelerate their transportation.

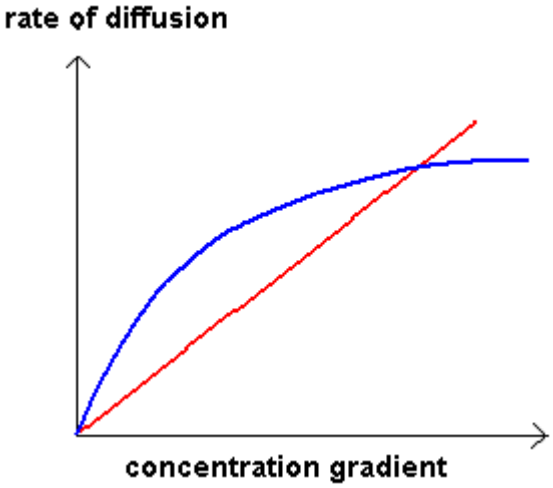


Fig. 4

7. Active transport through a membrane

It is intuitively clear that if some processes in living tissues are brought about in directions of diminishing concentration or under influence of electrostatic field, there also should be some processes in reverse directions. These processes need some external energy for transportation of ions or molecules against electric field or from lower to higher concentration. Such processes are called **active transport** and they use energy released in ATF hydrolysis or in light absorption. Active transport function in membranes is brought about by specialized protein carriers called **ion pumps**. Na^+/K^+ pumps provide higher concentration of potassium within a cell and lower concentration of sodium compared with ambient medium. H^+/K^+ pumps maintain concentration of H^+ ions in stomachs, while Ca^{2+} pumps – in muscles.

8. Action potential in a membrane

External stimuli can influence a state of membrane. Cell's reaction leads to change of the membrane potential. Excitation leads first to opening of sodium channels, as a result Na^+ ions due to simple diffusion penetrate the cell from the ambient medium and change to positive value electric

potential of inner side of the membrane with respect to outer side. As membrane potential changes, sodium channels are closing, while potassium ones are opening, therefore a part of K^+-ions leave the cell (simple diffusion) and membrane voltage is restored closer to the value of resting potential. Then refraction time comes and the membrane returns to the resting state, Na^+ / K^+ pumps restore normal concentrations of potassium and sodium ions in the cell. During refraction time the cell is unable to perceive new stimuli.

LECTURE 8. WAVE OPTICS

1. Light as an electromagnetic wave

Optics had developed apart from theory of electricity for centuries. Though many practically important devices like mirrors, glasses, terrestial telescopes had been used for centuries. J.C.Maxwell had formulated equations of classical electrodynamics (1864) and he had found that electromagnetic (EM) waves propagate with the speed of light. So it had been assumed that visible light is a kind of EM waves. Some years later, in 1887-1888 Heinrich Hertz proved that hypothesis experimentally. Hertz stated that EM waves with the wavelength of some meters could pass through dielectric media and reflect from metallic ones. Using metallic mirrors he changed direction of propagation of such waves and focused them. At the very end of XIX century Peter Lebedev observed birefringence of millimeter EM waves (1895), a phenomenon typical to propagation of light in some crystals (to be considered later in this Lecture). Lebedev even designed an analogue of Nicol prism for EM waves: two pieces of crystalline sulfur specially cut in appropriate direction and glued with ebonite.

Nowadays we consider visible light as a very short interval at the graph of possible wavelengths of frequencies of EM waves (*Fig. 1*).

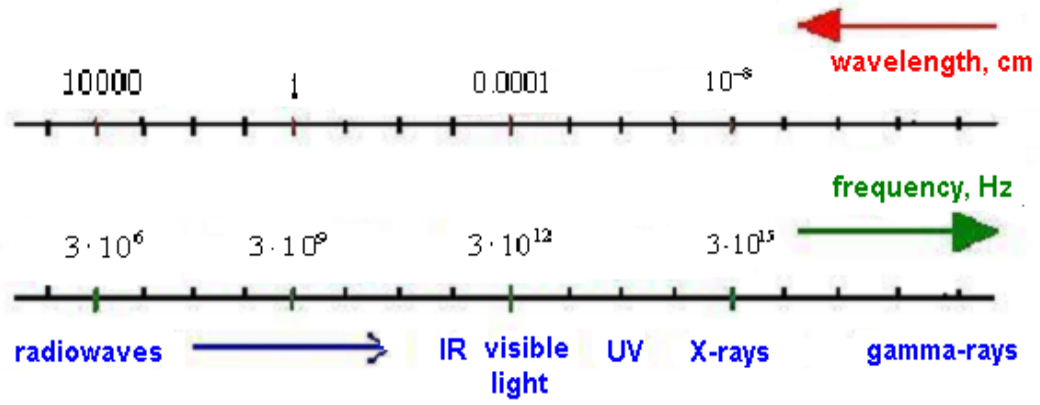


Fig. 1

Wavelengths of a visible light belong to a very short interval between 400 and 760 *nanometer*, although all conclusions of this lecture are also valid in the infrared and ultraviolet ranges of EM spectrum.

Light waves have transverse polarization, it means that the planes of oscillation of intensity of electric filed $\vec{E}$ and of magnetic induction $\vec{B}$ are perpendicular to each other and they both are perpendicular to the direction

of propagation (the direction of a light ray), coinciding with the vector of light speed in air (*Fig. 2*).

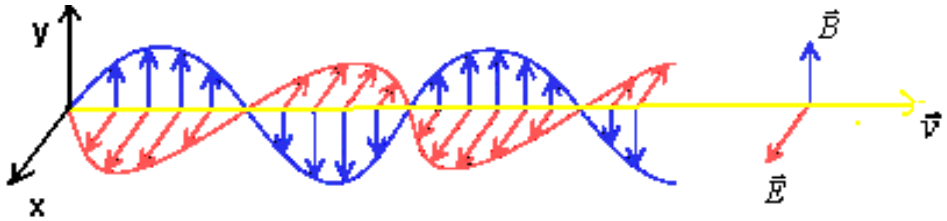


Fig. 2

Later in this and the next lectures we will consider only the vector of intensity (strength) of electric field $\vec{E}$, though the induction $\vec{B}$ is always present in EM waves. But the latter does not influence on human eyes, therefore traditionally it is not mentioned in optics.

2. Interference of light. Interferometry.

Interference is a phenomena, which is character for interaction of any kind of waves – mechanic, electromagnetic and, as we know now, gravitational too. Usually it is studied in the course of optics, because of some evident and practically important examples of this phenomenon observed in visible light. Mathematical description of interference we had considered previously in the Lecture 2, when being studied summation of two waves of the same frequency. Let us apply the same approach to interference in optics. Suppose the point *C* at a screen is illuminated with light of the same frequency (of the same color) generated by two sources *A* and *B* (*Fig. 3*).

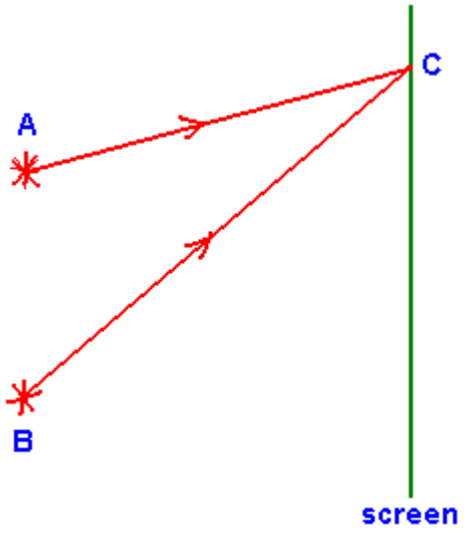


Fig. 3

Electric field strength in those two beams is equal to

$$E_1 = E_{01}\cos(\omega t + \varphi_1) \qquad \text{and} \qquad E_2 = E_{02}\cos(\omega\ t + \varphi_2)$$

Then sum of the fields is

$$E = E_0\cos(\omega\ t + \varphi), \tag{1}$$

where

$$E_0 = \sqrt{E_{01}^2 + E_{02}^2 + 2E_{01}E_{02}\cos(\varphi_1 - \varphi_2)}, \operatorname{tg}\varphi = \frac{E_{01}\sin\varphi_1 + E_{02}\sin\varphi_2}{E_{01}\cos\varphi_1 + E_{02}\cos\varphi_2} \tag{2}$$

In the theory of mechanical waves we did not care about polarization of waves because it had been unambiguously defined: acoustic waves in air or water always have longitudinal polarization. In optics situation is different: vectors $\vec{E}_1$ and $\vec{E}_2$ can be directed differently at the point *C*. Besides that as we will know from the next lectures natural light is generated in spontaneous transitions of atoms from some excited energy levels to lower ones. Spontaneity of those transitions means that they take place in random moments of time, direction of radiation and polarization of every portion (quantum) of emitted light are also random and the length of any quantum (it can be imagined as a length of a part of graph of *cos* function), therefore if an observer looks at the point *C*, phase difference $\varphi_1 - \varphi_2$ would be different at different moments of time. Average value of *cos* function is zero, therefore in general case amplitude is

$$E_0 = \sqrt{E_{01}^2 + E_{02}^2}$$

Absolute majority of light detectors (including human eyes) register not amplitude of EM wave, but intensity, which is proportional to the square of amplitude ($I \sim E^2$), so the previous formula can be rewritten as

$$I = I_1 + I_2 \tag{3}$$

Formula (3) gives the law of summation of intensities of **incoherent** light waves. Two different sources of natural light are always incoherent. On the contrary, **coherent** light sources generate rays with the same frequency, polarization and constant phase shift, so that the last, so-called interference term $2E_{01}E_{02}\cos(\varphi_1 - \varphi_2)$ in (2) should be taken into account.

A possible way of making of two coherent light sources is drawn at *Fig. 4*. It was proposed by Lloyd in XIX century. In fact, he used just one source of light at a point *A* and a mirror, so that light arrived at screen was sum of two beams: one passed directly and another one reflected from the mirror. A light source *B* in *Fig. 4* is the image of the source *A* in the mirror.

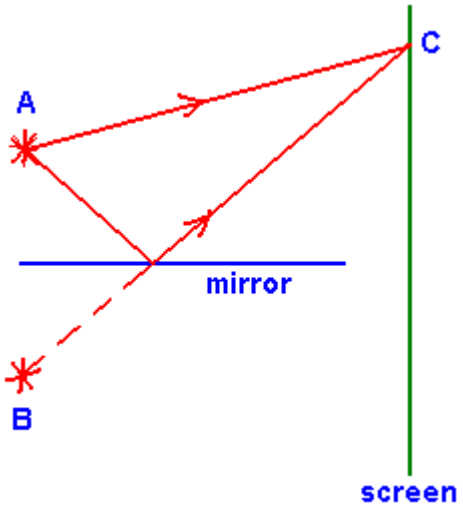


Fig. 4

It is seen from the scheme at *Fig. 4* that initial phases for both rays are the same, but some phase difference takes place, because reflected and passed directly rays propagate along different distances, they use different time to reach the screen. Evidently this phase difference is

$$\Delta\varphi = \omega(t_2 - t_1) = \omega\frac{l_2 - l_1}{c} \qquad (4)$$

Supposing that experiment at *Fig. 4* is carried out in air, we denote light speed in (4) with the letter c - like in vacuum, because in air refractive index (ratio of velocity of light in vacuum and in a given medium) is

$$n_{air} = \frac{c}{v_{air}} \approx 1$$

It means that as a result of interference (summation of two coherent beams) at the point *C* of the screen light intensity at that point is equal to

$$I = I_1 + I_2 + \sqrt{2I_1 I_2 \cos\Delta\varphi} \qquad (5)$$

Depending on value of phase difference there could be maximal, minimal or some intermediate brightness of light at the point *C*. It is clear

that there would be some difference of passed paths for both beams at any point adjacent with *C*, therefore intensity of light is different at different places – so the consequence of dark and bright points or curves is formed at the screen. This pattern is called interference one.

Interference of light is used for determination of concentration of gases or liquids in devices called **interferometers**. A scheme of the Jamin's interferometer is drawn at *Fig. 5* as an example.

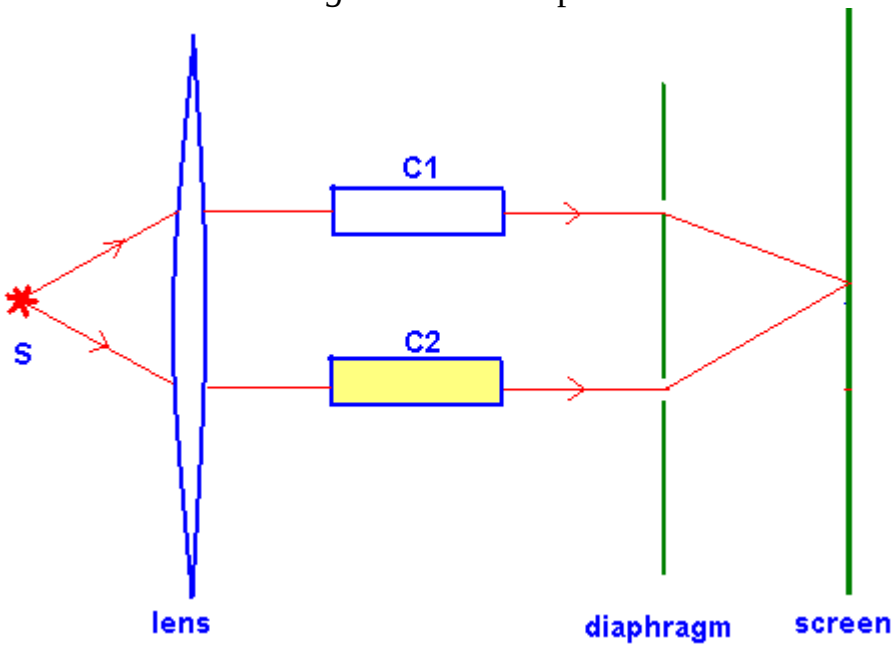


Fig. 5

Light beam from a source S is placed in the focal point of a converging lens. Two parallel beams after lens pass through two cells of equal size and then, after diaphragm, they form an interference pattern at a screen. If both cells are filled with same substance (air, for example), at the center of the interference pattern would be maximum of intensity. But if, for instance, the cell C_2 contains some other substance (even transparent), maximum of the interference pattern would be shifted, because of phase difference between two beams:

$$\Delta\varphi = \omega(t_2 - t_1) = \omega\left(\frac{l}{v_1} - \frac{l}{v_2}\right) = \frac{\omega l}{c} \cdot (n_1 - n_2)$$

3. Interference in thin films

Reflection of a light beam from a thin film represents an interesting and widely observed example of interference. Suppose a light beam is

incident at a thin transparent plate of thickness d (*Fig. 6*). An observer would detect a ray reflected from the upper side of the plate (ray 1 at the *Fig. 6*) as well as from the bottom side (ray 2). Suppose for simplicity that light is incident from air and that refractive index of a film is equal to n.

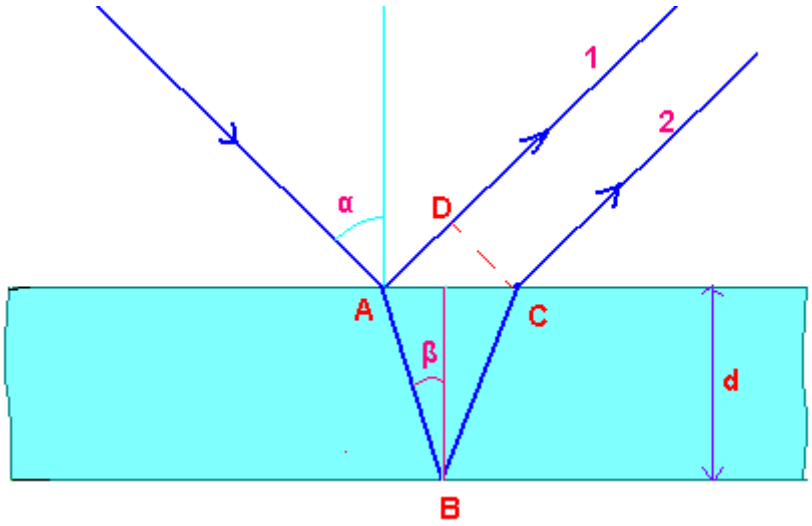


Fig. 6

Path difference between rays can be easily calculated

$$\Delta = 2 \cdot AB - AD = 2 \cdot \frac{d \cdot n}{\cos\beta} - 2 \cdot d \cdot \tan\beta \cdot \sin\alpha$$

Refractive index of the film substance is added in the numerator of the first term on the right-hand side of the last equation, because velocity of light propagation in the film is n times less than in air. As, according to the Snell's law of refraction $n \cdot \sin\beta = \sin\alpha$, the latter expression can be rewritten in a more useful form:

$$\Delta = 2 \cdot d \cdot \sqrt{n^2 - \sin^2\alpha} \qquad (6)$$

It was found experimentally that if a medium below the film has refractive index exceeding that of the film (for example, a gasoline film on water), then formula (6) is in excellent agreement with experimental results. In the opposite case (a soap film in air), phase of reflective beam gets addition equal π, so expression (6) is changed to

$$\Delta = 2 \cdot d \cdot \sqrt{n^2 - \sin^2\alpha} + \frac{\lambda}{2} \qquad (6, a)$$

here in (6, a) λ is the wavelength of the incident light beam.

If path difference is equal to even number of half-lengths (two waves are **in phase** with each other), there is a maximum of intensity of light in the point of observation:

$$\Delta = 2 \cdot m \cdot \frac{\lambda}{2}; \qquad m = 0,1,2,... \tag{7}$$

In the opposite case, if path difference is equal to odd number of half-lengths (he waves are **in antiphase**), there is a minimum of intensity of light in the point of observation:

$$\Delta = (2 \cdot m + 1) \cdot \frac{\lambda}{2}; \qquad m = 0,1,2,... \tag{8}$$

Iridescent colors of soap films, of oil films on water, of some insects' eyes are explained by light interference in films. This phenomenon is also widely used in optics for blooming of different devices like microscopes, photo cameras and binoculars – reduction of light reflected from lenses is brought about by use of very thin films coating lenses with thickness corresponding to odd number of wavelengths of green light, to which human eye is especially sensitive.

4. Diffraction of light. Huygens principle.

Light **diffraction** is any deflection from rectilinear propagation of light, which could not be explained as a result of reflection, refraction or curvature of a propagation path in media because of gradually changing refractive index. Like interference, diffraction is inherent in all kinds of waves.

It was Christiaan **Huygens** who first formulated a **principle** of drawing of fronts of diffracted waves. He stated that *every point in space being reached by wave motion becomes a source of secondary waves. The envelope of these secondary waves forms a wave front at the moment*. *Fig. 7* below (downloaded from the Wikipedia) illustrates the Huygens principle.

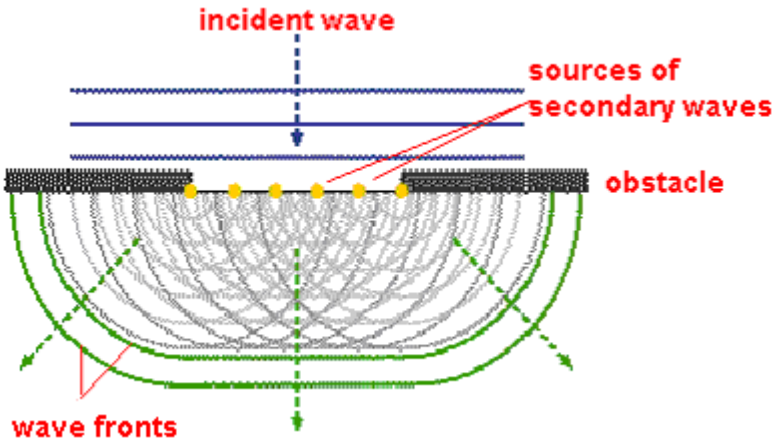


Fig. 7

5. Fresnel reformulation to Huygens principle. Fresnel zones.

Huygens principle qualitatively well enough explains diffraction, but it cannot help to calculate amplitudes of diffracted waves. A French physicist and mathematician Augustin-Jean Fresnel has made the next assumption (1821): wave motion in any part of space is a result of *interference* of secondary waves. By the way, this is a qualitative explanation why wave at *Fig.7* propagates only in one direction from the slit in the screen and there is no reversed wavefront. According to Fresnel's explanation, backward secondary wave, when interfering with continuous incident wave, is extinguished.

At the end of XIX century Gustav Kirchhoff proposed a method of calculation of amplitudes of secondary waves. He supposed that if a point S is a source of light wave (*Fig. 8*), then at some moment of time that wave reaches a sphere of radius r_0. Amplitude of electric field in some point of observation P would be the result of summation of waves generated in all points placed at this sphere.

For example, amplitude of a wave generated by a small part of the sphere near a point *A* is equal to

$$dE_A(P) = \frac{E_0 \cos(\omega t - kr)}{r} \cdot K(\psi) \cdot d\Omega \qquad (9)$$

where E_0 is the amplitude of a light wave at the point *A*, ω and $k = \omega/c$ are the cyclic frequency and the wavenumber of the wave respectively, $d\Omega$ is surface of the small part of the sphere around *A*, $K(\psi)$ is some numerical

coefficient depending only on the angle ψ between a normal to the sphere at *A* and the straight line connecting points *A* and *P*.

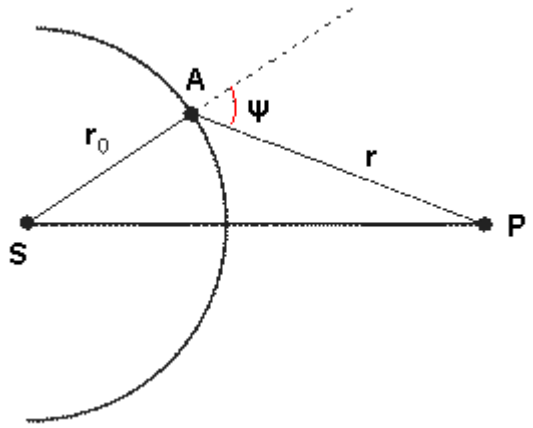


Fig. 8

It had been proven that $K(\psi) \sim 1 + \cos\psi$. Symmetry considerations allows to realize that all points on the mentioned sphere remote from the point of observation at the same distance as the point *A* make the same contribution into interference. All such points belong to an arc, which could be drawn at the surface of the sphere by a circle of radius *r* and the center in *P*.

This helps to understand the concept of **Fresnel zones** – the geometrical construction, shown at *Fig. 9*.

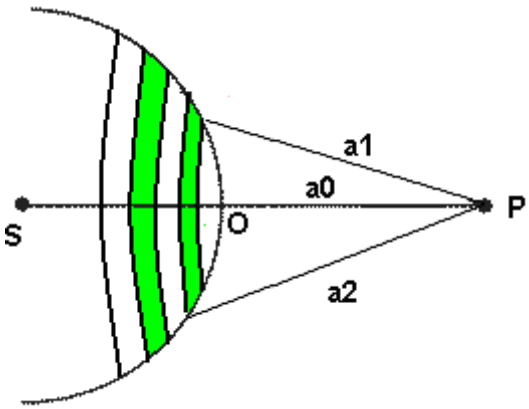


Fig. 9

Consider the spherical wavefront a point *O*, which is at the shortest distance between *S* and the point of observation *P*. Suppose, the distance *OP* is a_0. Then we draw at the surface of the wavefront a circle so that distance between any point at this circle and the observation point *P* is equal to

$$a_1 = a_0 + \frac{\lambda}{2} \tag{10}$$

where λ is the wavelength of light at frequency ω: $\lambda = 2\pi c/\omega$. The round set of points remote from *P* at distances more than a_0 and not more than a_1 is called the **first Fresnel zone** for the light source *S* and the point of observervation *P*.

Similarily we can draw at the sphere circles of the next Fresnel zones (the second, the third, the forth and so on), only first five of them are shown in *Fig. 9*.

$$a_2 = a_1 + \frac{\lambda}{2}, \quad a_3 = a_2 + \frac{\lambda}{2}, \quad a_4 = a_3 + \frac{\lambda}{2}, ..$$

It evidently follows from geometry that interference of light waves generated in two equal small areas $d\Omega$ of neighbor zones leads to some extinguishing of amplitude of resultant light, because due to condition (10) those waves would be in antiphase (recall formula (8)) at every moment of time. Therefore the amplitude of a secondary wave generated by every even ring at *Fig. 9* approximately compensates amplitude of light generated at previous odd ring (or circle – for the first of them), that is neighboring terms of the next series have opposite signs:

$$E(P) = E_1 + E_2 + E_3 + E_4 + ... + E_N \tag{11}$$

where (integration is made on the surface of the central circle for E_1 and on the area of the *i*-th ring for E_i)

$$E_i(P) = \int \frac{E_0 \cos(\omega t - kr_i)}{r_i} \cdot K(\psi_i) \cdot d\Omega$$

Nevertheless resultant amplitude $E(P)$ is not zero in (11) because of gradual slight change of coefficients $K(\psi_i)$ and distances r_i from a zone to zone. It should be mentioned here that thickness of every ring is grossly exaggerated at *Fig. 9*, while distances *SO* and *OP* are significantly underestimated. Total number of Fresnel zones increases with increase of radius *SO* and distance *OP* and decreases with increase of a wavelength. It could be estimated that for visible light and $SO = OP = 1$ *m* radius of the first Fresnel zone is about 1.1 *mm*. It is evident that

$$|E_1| > |E_2| > |E_3| > |E_4| > \ldots > |E_N| \qquad (11, a)$$

therefore we can exclude from equation (11) the first and the last terms in approximate estimations, so that the residual would be approximately the same in value, but of opposite sign (we remember that terms in every couple of neighboring terms in (11) are in antiphase to each other):

$$E(P) \approx -E_2 - E_3 - E_4 - \ldots - E_{N-1} \qquad (12)$$

The more is N, the more accurate is expression (12). Adding left-hand parts and right-hand parts of equation (11) and (12) we obtain

$$E(P) \approx \frac{E_1 + E_N}{2}$$

Naturally $|E_N| << |E_1|$, so we can conclude that

$$E(P) \approx \frac{E_1}{2} \qquad (13)$$

In other words, amplitude of intensity of electric field of a light wave in the point of observation is about half of amplitude generated by the first Fresnel zone. This interesting fact is completely confirmed experimentally: placing at the point O of an opaque screen perpendicular to the direction SP with a small round hole leaving open only the first Fresnel zone leads to the 4-fold increase of light intensity at the point P - in accordance with (13).

Similarly, if we place at the point O a small round disk, which overlaps completely some first Fresnel zones, we have to get a bright spot in the point P, because removal of first terms from (11) changes a whole sum a little. This phenomenon was called the **Poisson spot** after a French physicist Poisson, who had first payed attention to this consequence of Fresnel theory and used this fact as an argument against the zone theory. It was very surprising when another one physicist Arago experimentally observed such spot. Therefore sometimes this spot is named after Arago, or after Fresnel.

Fresnel's zone theory does not contradict evident rectilinear light propagation in uniform media. Indeed, if the plane of a screen intersects line SP at the point O (*Fig. 9*), so that only some first zones are open, then intensity of light in the point P would be about the same as without the screen because of condition (11, a). On the contrary, if the screen at O

closes sufficient number of zones then light intensity in the point of observation tends to zero (in the above mentioned Arago's experiment increase of radius of an obstacle had led to decrease of light intensity in the spot).

6. Light diffraction on a slit

Let us now consider another one application of Huygens-Fresnel principle.

Suppose a parallel beam of a monochromatic light is incident at a slit in an opaque screen (*Fig. 10*). The width of the slit is equal to b – a value of the order of the wavelength of light λ. The length of the slit in the direction perpendicular to the plane of the picture is much more than its width, so that for simplicity we can believe this length to be infinitely long. Every point of aperture is a source of secondary waves, which can propagate in every direction behind the screen, although amplitudes of waves in those directions would be result of interference. We consider light beam diffracted by the slit at the angle φ to the direction of incident beam. Light from this beam can be collected with a lens. It is easily seen that path difference for two rays diffracted at the opposite sides of the slit is equal to

$$\Delta = b \cdot \sin\varphi$$

If this path difference is equal to even number of half-waves, then waves, which come to lens from different parts of the slit aperture, extinguish each other, so the condition of minimal light amplitude is

$$b \cdot \sin\varphi = 2 \cdot m \cdot \frac{\lambda}{2}; \qquad m = 0,1,2,... \tag{14}$$

while the maximum takes place if

$$b \cdot \sin\varphi = (2 \cdot m + 1) \cdot \frac{\lambda}{2}; \qquad m = 0,1,2,... \tag{15}$$

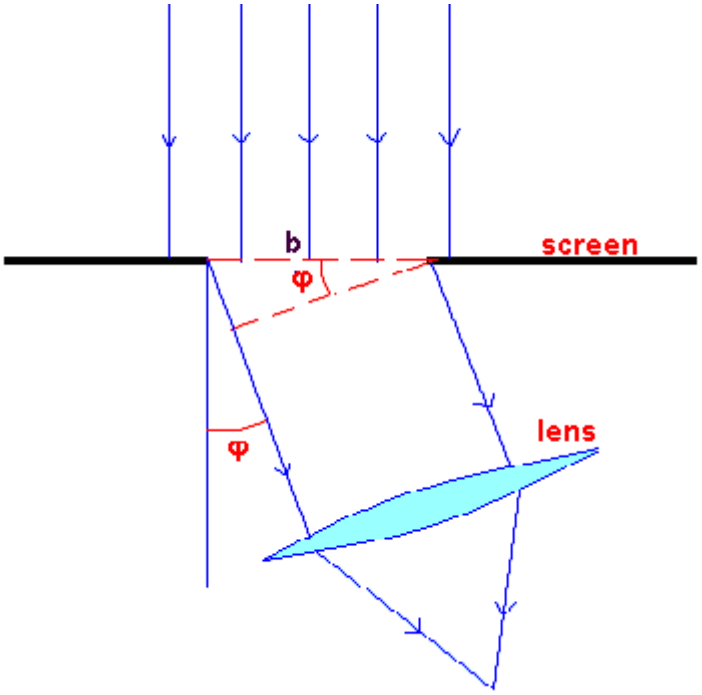


Fig. 10

7. Diffraction grating

Diffraction grating is a glass or metallic plate, at which a periodic structure of parallel grooves or slits is made with the help of a special machine. Grooves are very narrow: there can be up to thousands slits per millimeter of the plate. (Only four grooves are drawn at *Fig. 11*).

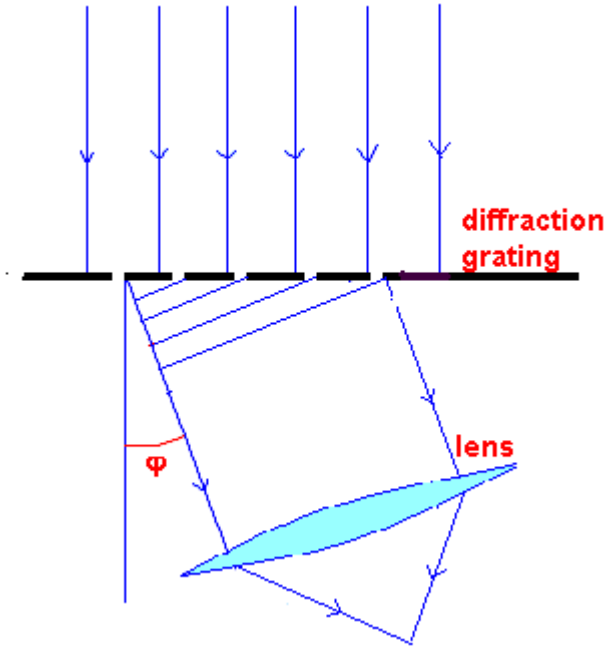


Fig. 11

It is seen from *Fig. 11* that light beams passed through (or reflected from) different grooves of a diffraction grating would be added in phase if

$$d\sin\varphi = 2m\frac{\lambda}{2} = m\lambda; \qquad m = 0,1,2,... \tag{16}$$

where $d = a + b$ is the period of the grating, b is the groove's width and a is the space between neighboring grooves. Formula (16) is the condition of **principal maxima** of a diffraction grating. A whole number m in this case is called the **order of a maximum**. Central maximum has order $m = 0$. Between every couple of principal maxima $N - 1$ minima take place. Derivation of condition for those minima is a bit more complicated therefore we give only the final result (number of groves in a grating is denoted by N):

$$d\sin\varphi = \left(m + \frac{l}{N}\right)\lambda; \qquad m = 0,1,2,...; \qquad l = 1,2,...N-1 \tag{17}$$

As any diffraction grating consists of slits and every slit does not radiate light in the direction given by then the condition (14), the condition of principal minima of a grating obeys the equation:

$$b\sin\varphi = (2m+1)\frac{\lambda}{2}; \qquad m = 0,1,2,... \tag{18}$$

It is possible that in some directions, given by (16) the condition (18) is satisfied too. In those cases an appropriate maximum does not appear.

Conditions of **subsidiary minima** of a grating follows from (16)

$$d\sin\varphi = (2m+1)\frac{\lambda}{2} = \left(m + \frac{1}{2}\right)\lambda; \qquad m = 0,1,2,... \tag{19}$$

There is a **subsidiary maximum** between every two subsidiary minima. It is found that between two adjacent principal maxima there are $(N-2)$ subsidiary maxima, if a grating contains N grooves. Though intensity of light in the brightest of subsidiary maxima does exceed some percent of light intensity in any of principal maxima.

Diffraction gratings play independent role in chemical analysis, as we will know in the next part of the lecture. Besides that, they same mathematical approach is used in x-ray determination of structure of

extremely small periodic objects like crystal lattices and DNA chains. This topic will be also considered in the Lecture 12.

8. Diffraction grating as a spectral device

Light spectrum is distribution of a light beam on frequencies or wavelengths. In fact, the spectrum of electromagnetic waves is drawn at *Fig. 1*, it contains visible light as a narrow interval of a semi-infinite graph. When considering diffraction phenomena we have considered only monochromatic light: the cyclic frequency ω (or the wavelength λ) was constant everywhere in our reasoning until now. But since Newton we know that the "white" sunlight is a composition of many different colors (many different frequencies/wavelengths). Different sound beams can contain light of different frequencies with different intensities, depending on chemical composition of a light source. Therefore study of spectra of light from different sources (called **spectral analysis** or **spectroscopy**) is a useful tool of chemical analysis.

The fact that diffraction grating can be used as a spectral device is easily seen from formulas (16)–(19): they all contain wavelength. It means that angular position (φ in expressions (16)–(19)) of every maximum or minimum depends on the wavelength (or, in other words, on the color) of light. Only the central (zero-order maximum) remains white, if daylight is incident at a diffraction grating, all other maxima are colored and do not coincide for different colors.

To distinguish quality of different spectral devices a notion of the **angular dispersion** is introduced. This parameter states a measure of angular distance between different wavelengths: $d\varphi/d\lambda$. Differentiating both sides of Equation (16) we can determine value of angular dispersion of a diffraction grating with period d:

$$d\left(d\cdot\sin\varphi\right)=d\left(m\cdot\lambda\right)\ \Rightarrow\ d\cdot\cos\varphi\cdot d\varphi=m\cdot d\lambda$$

and finally

$$\frac{d\varphi}{d\lambda}=\frac{m}{d\cos\varphi} \qquad (20)$$

It is seen from (20) that better resolution corresponds to a smaller period of a grating and to a higher order of maximum (m is in numerator, $\cos\varphi$ is in denominator).

If observation is carried out at small diffraction angles $\varphi \approx 0$, then formula (20) becomes simpler:

$$\frac{d\varphi}{d\lambda} \approx \frac{m}{d} \qquad (20, a)$$

Nevertheless, increase of angular dispersion does not mean that two rather close spectral lines (let's denote their wavelengths as λ and $\lambda_1 = \lambda + d\lambda$) are perceived as separate lines. Situation is complicated by indispensable presence of some minima and subsidiary maxima around every principal maximum of a given color. The less wavelength difference between neighbor principal maxima for different colors, at which they are considered as different lines, is called **spectral resolution**. Spectral resolution of a diffraction grating is usually estimated with value called **resolving power** *R*:

$$R = \frac{\lambda}{d\lambda} \qquad (21)$$

Lord Rayleigh proposed the next criterion of determination of the resolving power of a grating, later named after him: suppose we watch two spectral lines of different colors (*Fig. 12*):

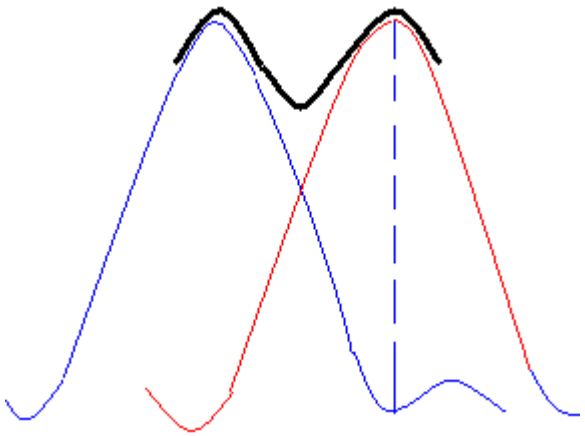


Fig. 12

We consider them as different if maximum of one line coincides with the nearest minimum of another one. Then we believe that they are resolved. Using conditions (16) and (17) we can write:

$$m\lambda_1 = \left(m + \frac{1}{N} \right) \lambda$$

Hence, after simple algebra we have the next expression for the resolution of a diffraction grating:

$$R = mN$$

The more is number of grooves in a grating the better its spectral resolution.

Resonableness of the Rayleigh criterion had been confirmed experimentally. Indeed, as peaks at *Fig. 12* had different frequencies they could not interfere, therefore their intensities were added in compliance with Eq. (3). It was counted that light intensity in the saddle point between two peaks at *Fig. 12* was about 80% of maximal value. As it have been stated experimentally it is sufficient for a human eye to consider those peaks as different lines.

LECTURE 9. MOLECULAR OPTICS

In the previous lecture we considered some important optical phenomena based on the wave nature of light. The present lecture is devoted to some special features of light propagated through matter.

1. Plane-polarized light

We know that light like any other electromagnetic wave has transverse polarization. Any beam of natural light contains many photons (which can be imagined as fragments of sinusoidal waves) with different polarization. Therefore in every rather intense portion of a natural light wave we can find equally probable all directions of oscillation of electric field vector $\vec{E}$. But it was found that sometimes such non-polarized wave can be transformed into so-called plane-polarized one. Let us consider light refraction at the boundary between two dielectric media (*Fig. 1*):

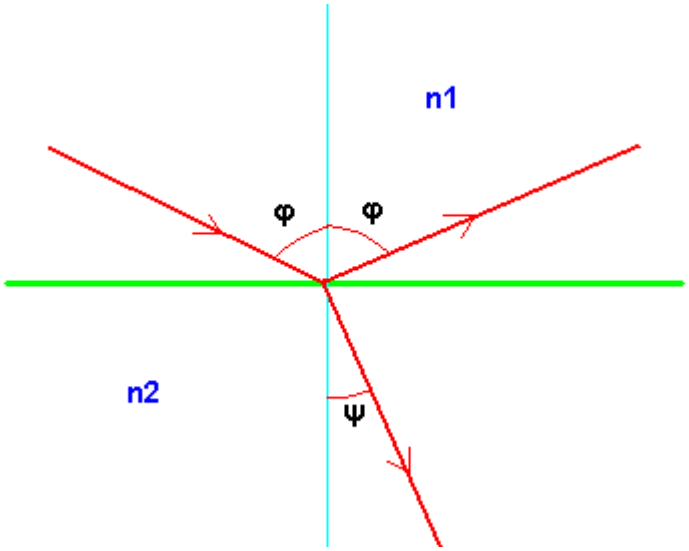


Fig. 1

If a non-polarized light is incident at the boundary so that angle between reflected and refracted beams is equal 90^0, then reflected light is completely **plane-polarized** (or in other words **linearly polarized**) – that is vector $\vec{E}$ oscillates in one plane. It would be the plane perpendicular to the plane of picture. (Refracted light remains non-polarized.) It can be easily derived from the Snell's law of refraction that angle of incidence in this case obeys the formula (try to derive it):

$$\tan \varphi_B = \frac{n_2}{n_1} = \frac{v_1}{v_2} \qquad (1)$$

where n_1 and n_2 are refractive indices in the first and the second medium respectively, v_1 and v_2 are correspondent light velocities. Angle φ_B is called the Brewster angle, after David Brewster who found experimentally in 1815 that linearly polarized in the plane of incidence light is completely refracted (there is no reflected beam), if it incident at the angle φ_B. This is a consequence (and a confirmation) of the fact that EM waves are transverse, because it is evident that if light incident at the Brewster is polarized in the plane of incidence, then reflected beam should be polarized in the direction of its propagation, but this is impossible.

Light reflection at the Brewster's angle is a simple example of a way of preparation of a plane-polarized light beam, though it does not have practical significance, as intensity of a reflected beam is very small. Alexander Stoletov tried to save the day using a set of plates (Stoletov's pile) to increase intensity of a plane-polarized light component, but this rather bulky way is not used nowadays.

2. Birefringence and polarizers

More widely spread and convenient way of preparation of a plane-polarized light uses phenomenon of birefringence in crystals. Anisotropy of physical properties is typical for crystals. As for the light refraction, it was found that all but cubic crystals have one or two so-called optical axes – these are directions, in which light propagates like in isotropic media. (Cubic crystals are completely isotropic for light propagation.) For simplicity we consider here only optically uniaxial crystals, all the more they are widely used. If a beam of natural (non-polarized) light passes in a crystal not parallel to the optical axis, then it splites into two rays, both linearly polarized. One ray has the same phase velocity as light propagating parallel the optical axis. Its polarization plane is parallel to the direction of propagation and to the optical axis, this beam is called the **ordinary** one. Another beam, called **extraordinary**, propagates in some other direction, its polarization plane is perpendicular to that of the ordinary ray, and its speed and the refractive index n_e differ from values of the ordinary beam.

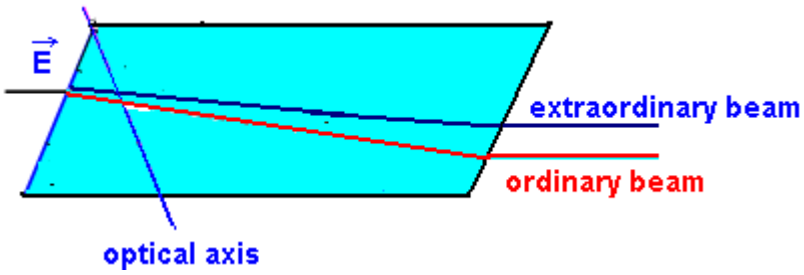


Fig. 2

Natural light polarized by birefringence phenomenon in a crystal of Iceland spar is schematically drawn at *Fig. 2*. Incident beam (from the left) is non-polarized, ordinary beam is polarized in the plane of the picture, extraordinary beam is perpendicular to this plane. Birenfingence was first discovered by a Dutch scientist Bartolinius in 1669 right in Iceland spar. Later Huygens proposed the first explaination of this phenomenon and found that both resultant rays are linearly polarized.

The advantage of the Iceland spar crystal is in rather big difference in refraction indices: for yellow light $(n_o - n_e)/n_o$ can achieve 10%. William Nicol proposed a very popular concept of a **polarizer** using two such crystals. It is used to prepare a plane-polarized beam from the natural light. Construction of the Nickol's prism is schematically drawn at *Fig. 3* (from the Wikipedia). Two crystals of Iceland spar are cut at special angles

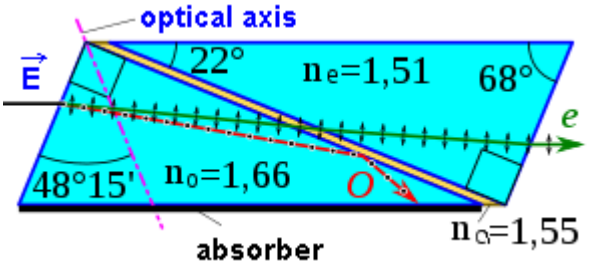


Fig. 3

with respect to their optical axes and clued with Canadian balsam, because its refractive index $n_e < n < n_o$. Therefore extraordinary beam *e* passes through the clued interface between crystals, while ordinary one *O* reflects from that interface (and further is absorbed). So the 100% plane-polarized extraordinary beam passes from the Nicol's prism in the sequel. It is clear from symmetry considerations that intensity of extraordinary beam is half of intensity of incident natural light.

If a linearly polarized light beam of intensity I_0 is incident at the Nicol's prism, but the electric field vector $\vec{E}$ constitutes some angle φ with

the polarization plane of the prism, then, as it was stated first by a French engineer Malus, intensity of the resultant beam obeys the next **Malus law**:

$$I = I_0 \cos^2 \varphi \qquad (2)$$

Problem. Try to derive this law, in fact it is necessary to decompose the vector $\vec{E}$ into two components: one coinciding with the prism polarization plane and another one perpendicular to it.

Malus law gives the way to determine direction the polarization plane of linearly polarized light with help of a polarizer, which is called an **analyzer** in this case.

It should be added that if a natural light beam is incident at the uniaxial crystal perpendicular to its optical axis then ordinary and extraordinary rays propagate in the same direction, but with different speed and mutually perpendicular polarization. Therefore output beam has some phase difference between light polarized parallel to the axis and perpendicular to it.

3. Rotation of polarization plane in optically active media. Biot law.

If one crosses polarization planes of two consecutive polarizers (they can be considered as the polarizer and the analyzer) they would not transmit light. But it was found that placing of a uniaxial crystal between them so that light should propagate parallel to its optical axis leads to appearance of some light after the analyzer. (Quartz crystal is often used for such demonstration.) It means that quartz changed (rotated) direction of linearly polarization of light propagating between polarizers. Rotation of analyzer's polarization plane shows that angle of that rotation is proportional to the crystal's thickness l:

$$\varphi = \alpha \cdot l \qquad (3)$$

α is the specific rotation of a crystal, its measured in *degree/sm* usually. It depends on nature of the crystal and of light wavelength.

Similar phenomenon was also registered in some liquids. But analogous formula for liquids (it is called **Biot law**) contains not only the thickness of a liquid layer l, through which light penetrates, but also the liquid's concentration c:

$$\varphi = \alpha \cdot c \cdot l \qquad (4)$$

Substances rotating polarization plane of a linearly polarized light beam are called **optically active**. This effect was explained by Fresnel, who assumed the every linearly polarized beam contains equal components, in one of which polarization the plane rotates clockwise, and in another - in opposite direction. In air or in vacuum those components compensate each other and plane of polarization does not change. But when passing through optically active media, they propagate with different velocities, therefore polarization plane of their sum is finally turned out to be rotated. An experiment on rotation of the polarization plane in an optically active liquid is schematically drawn at *Fig. 4* (from the Wikipedia).

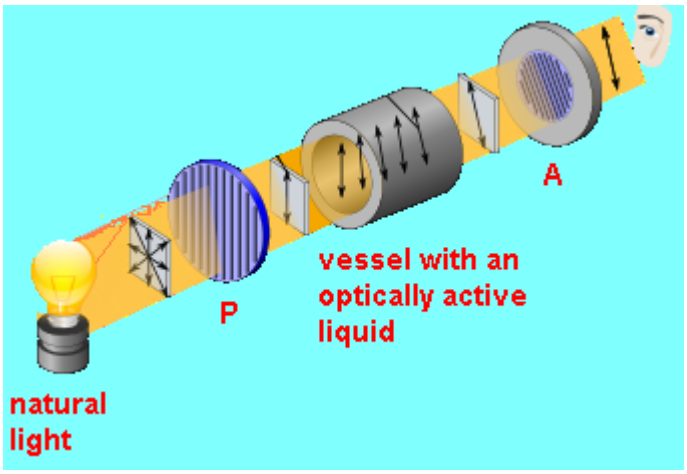


Fig. 4

Biot law is widely used in experiments on determination of both (so-called left-handed and right-handed) components of many liquids (water solutions of water, urine, acids, nicotine and many others). It is very important practical problem, because chemically grown substances usually contain equal amounts of both components, but sometimes one of them can be useful, but another one is dangerous. They are indistinguishable by chemical methods, but their optical and biological properties are different.

4. Dispersion of light

We have mentioned previously that refractive index of light depends on its wavelength (or frequency). This phenomenon is called **dispersion**, it is common for all media (and for all kinds of waves, in truth). One of the first experiments on light dispersion was carried out by Sir Isaac Newton, when he used a glass prism to resolve white light into components of different colors. He had found that red beam changed its direction of propagation less than all others, while violet – more than others. It followed

from that observation that refractive index of red light is minimal and of violet one is maximal of all visible colors. Classical theory of light dispersion in dielectrics was developed by Hendrick Antoon Lorentz. He had considered one electron in atom, which was moving under influence of electric field of all other atoms and electrons as well as under the influence of electric field of propagated through the medium light wave. (Later Lord Rayleigh had found in archives that J.C.Maxwell had been aware about the results discussed below, because he had proposed similar examination problem at the Cambridge university some years before the Lorentz's work). Using Lorentz's approach we can write a differential equation of electron's motion as the second Newton's law of dynamics:

$$m\frac{d^2x}{dt^2} = -kx - r\frac{dx}{dt} + q \cdot E_0 \cdot \cos\omega t$$

here in equation (5) m and q are the mass and the electric charge of the electron, k is some elastic modulus characterizing a restoring force acting at the electron from surrounding atoms, r is some coefficient characterizing attenuation, E_0 is the amplitude of electric field of the light wave, ω is the cyclic frequency of light. For simplicity we consider electron's motion in a direction of $x-$axis.

Dividing both parts of the latter equation by the mass of electron we have

$$\frac{d^2x}{dt^2} + 2\gamma\frac{dx}{dt} + \omega_0^2 x = \frac{q \cdot E_0}{m} \cdot \cos\omega t \qquad (5)$$

Equation (5) looks similarly to Eq. (6) from the second Lecture. From mathematical point of view this is the same equation, so we can use the previously derived solution just changing some variables:

$$x = \frac{E_0}{m} \cdot \frac{1}{\sqrt{\left(\omega_0^2 - \omega^2\right)^2 + 4\gamma^2\omega^2}} \qquad (6)$$

Frequency of light waves are very high, therefore electron oscillation is forced and its frequency is equal to frequency of the external field of EM wave:

$$x(t) = x_0 \cos\omega t = \frac{E_0}{m} \cdot \frac{\cos\omega t}{\sqrt{\left(\omega_0^2 - \omega^2\right)^2 + 4\gamma^2\omega^2}}$$

So every atom containing of a positive nucleus and a negative electron forms a pulsing electric dipole with the dipole moment equal to

$$p(t) = q \cdot x(t) = q \cdot x_0 \cdot \cos \omega t = p_0 \cdot \cos \omega t$$

Dipole moment of a unit volume P_0 of such dipoles is N times more, (N is the concentration of dipoles). It is derived in classical electrodynamics that dielectric permittivity of a non-conducting medium is equal $\varepsilon = 1 + 4\pi P_0 / E_0$ and refractive index is the square root of dielectric permittivity, so we have

$$n^2 = \varepsilon = 1 + \frac{4\pi q E_0}{m\sqrt{\left(\omega^2 - \omega_0^2\right)^2 + 4\gamma^2\omega^2}} \qquad (7)$$

At the appropriate frequency of an external force the resonance phenomenon takes place leading to rapid growth of amplitude of oscillations, as we know from the Lecture 2. It remains valid in the case of an electrically charged particle vibrating in the presence of a propagating EM wave.

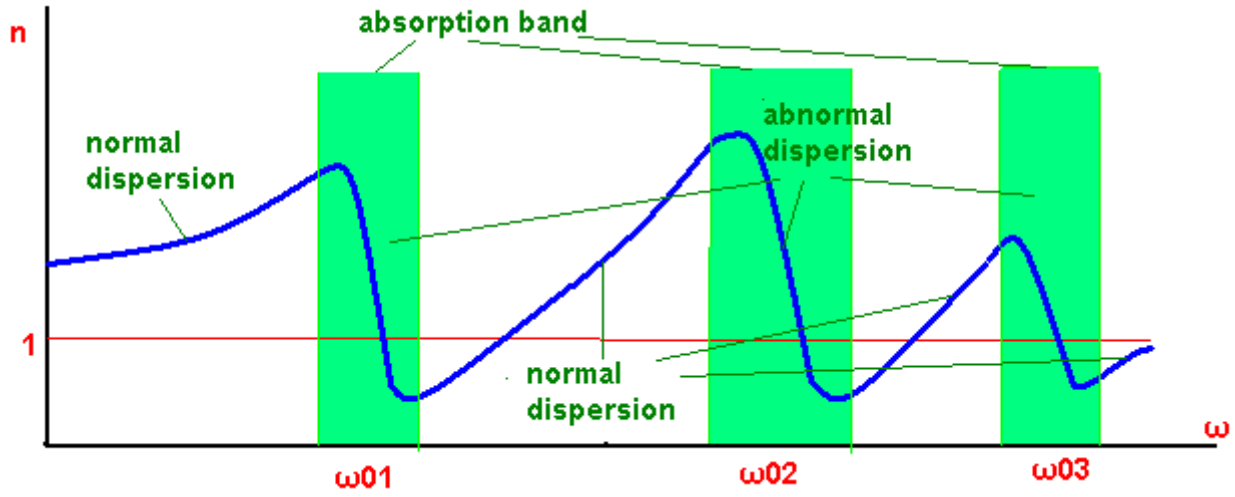


Fig. 5

But in optics situation is a bit more complicated, because real substances usually consist of different kinds of charged particles: along with electrons there could be different ions. Polar molecules can also take part in rotational motion under the influence of varying electric field. Therefore the denominator of formula (7) has several minimums, corresponding to

frequencies of inherent vibrations of different particles and graph of dependence of refractive index on the frequency has several maximums. Typical dependence of this kind is shown at *Fig. 5*. For law-frequency EM waves (radio frequency band) n slowly increases with increase of ω. This is the case of **normal dispersion** ($dn/d\omega > 0$), a phenomenon well-known since Newton's experiments with prisms. As ω comes up to ω_{01} (usually ω_{01} is the frequency of inherent oscillations of ions), amplitude of the refractive index reaches a local maximum, but dispersion becomes **abnormal (anomalous),** because $dn/d\omega < 0$. It was found that abnormal dispersion is always accompanied with absorption of external EM wave (absorbed energy is used for maintaining particles' forced oscillations). The first absorption band belongs to the infrared range for many substances. Thereupon situation is repeated: frequency range of normal dispersion and following absorption band in vicinity of electrons' inherent frequency of oscillations, usually this absorption band is in ultraviolet or visible part of spectrum (near ω_{02} at *Fig. 5*). Typical absorption band corresponding to fast forced rotation of dipole molecules usually is in the *X*-rays range of frequencies (ω_{03}). Refractive index of *X*-rays is slightly less than 1, as it is showed at the graph. Of course, this description is rather simplified: real substance could contain different kinds of ions having different mass and therefore different inherent frequencies of vibrations. Experiments with the short-wavelength radiation showed that the high-energy *X*-rays or gamma-radiation can cause oscillations of electrons of inner atomic shells, thus another one absorption band takes place far on the right part of spectrum.

Absorption of any kind of waves is usually introduced as an imaginary addition to the wavenumber. More thorough theory than that considered above states that refractive index of any substance has a complex value:

$$\sqrt{\varepsilon} = n(\omega) - i\kappa(\omega) \qquad (8)$$

where $\kappa(\omega)$ is the **attenuation coefficient** of a considered media. Cyclic frequency in both terms on the right-hand side of (8) emphasizes the fact that attenuation coefficient as well as refractive index depends on frequency. For a transparent substance, refractive index of which is similar to that drawn at *Fig. 5*, value of attenuation coefficient is close to zero in all ranges of normal dispersion and quickly grows to local maximums in every band of anomalous dispersion.

4. Spectroscopy. Bohr postulates

As Newton's prism because of dispersion deflects light of different color (different wavelength) at different angles, it was used to study spectra of different substances. It was found that heated solids emit **continuous spectra**, hot substances in gaseous state can also generate **discrete** and **band spectra**. It became clear to the beginning of XX century that discrete spectra consisting of narrow lines are emitted by gaseous atoms or ions, while band spectra composed of large amount closely spaced lines are radiated by molecules.

Emission spectra consist of vertical bright lines or bands at a black background (emission spectrum of atomic hydrogen is shown in *Fig. 6* (from the Wikipedia). **Absorption spectra** contain black lines or bands at the colored background.

Fig. 6

Every chemical element radiates and absorbs light at the same its own specific wavelengths, so the spectrum can be considered as "the optical ID" of an atom. It should be mentioned here that helium had been first discovered by astronomers in their investigations of spectrum of the Sun and only about 30 years later it was found at our planet.

Ernest Rutherford stated experimentally that structure of any atom resembled our Solar System – from semi-classical point of view light negatively charged electrons rotate along their orbits around heavy positive nucleus (1911). It means that at any moment of time every electron moves with acceleration, in other words some changing in time electric current flows continuously in every atom. According to Maxwells' electrodynamics electromagnetic waves have to be radiated by accelerated charges' motion, taking away energy. Finally electron should loose all its energy and fall on its nucleus. Nevertheless as many experimental observations show atoms do not perish for thousands and millions years.

To reconcile theory with experiment Niels Bohr formulated the next postulates in 1913:

1) *Every atom (or a system of atoms) can exist only in some selected states characterized by some* ***discrete*** *values of energy* $E_1, E_2, E_3, \ldots$ *These energy states are called* ***stationary****. Atom does not radiate electromagnetic waves in stationary states. (This postulate is also called the* ***rule of energy quantization****.)*

2) *If an atom passes from the state with higher value of energy* E_m *to a state with lower energy* E_n*, its energy diminishes by* $E_m - E_n$*. If this transition is accompanied with radiation, one quantum of electromagnetic radiation is emitted, its energy is* $hf = E_m - E_n$*. (Radiationless transitions are also possible.) Transition in opposite direction needs some external energy, which can be obtained in absorption of a quantum.*

Values *f* and *h* are the frequency of a quantum of electromagnetic radiation and the Planck constant respectively. They will be discussed more minutely in the next Lecture.

6. Light scattering

Light **scattering** is change of direction of propagation of light because of its interaction with molecules of a substance. John Tyndall was first who studied experimentally light dispersal in turbid media. He had found that white light seemed to be sky-blue, when passing through a gas or a liquid containing suspended small particles of impurities. Lord Rayleigh proposed theoretical explanation of this phenomenon based on equations of Maxwell electrodynamics. Rayleigh considered the case of rather long light wavelengths compared with sizes of particles of impurities. He managed to derive an expression for intensity of light dispersed by such particles. This intensity turned out to be inversely proportional to the forth power of the light wavelength:

$$I_{sc} = \frac{I_0}{\lambda^4} \cdot \frac{\left(1+\cos^2\theta\right)}{r^2} \cdot \frac{\pi^4 d^6}{8} \cdot \left(\frac{n^2-1}{n^2+2}\right)^2 \qquad (9)$$

where I_0 is intensity of incident light, θ is the angle between directions of propagation of the incident and of the scattered beams, *r* is a distance from the scattering volume to observer, d is the average diameter of particles scattering light.

This result explained the coloring of white light in Tyndall's experiments, as well as the blue color of a sky in a sunny day. Indeed, cosmonauts observe the Sun and bright stars against a black background beyond the Earth's atmosphere, because light from the Sun and stars does not disperse in cosmic space. The Earth's atmosphere, containing some amount of dust, compels any light beam passing to our eyes from the Sun to change its direction many times because of dispersion, therefore all sky is bright in daytime. Blue light waves have about twice shorter wavelength than red light, therefore, even if their initial intensity was equal, intensity of

scattered blue light is 16 times more than of red one according to (9). The same formula also explains reddish colors of sky in sunrise or sunset: scattered light with longer wavelengths propagates mostly in the direction of incident light.

It seemed that intensity of scattered light should diminish with altitude, because of diminishing of dust concentration, which obeys Boltzmann's distribution in the gravitational field. However it was found that scattering becomes stronger (sky is more intensively colored in mountains). Rayleigh assumed that this kind of scattering caused by air molecules. Later Marian Smołuchwsky explained this phenomenon by random temperature fluctuations of refractive index of air (1908). Though Rayleigh did not know about this mechanism, he developed a theory based on consideration of random changes of refractive index in very small, but containing a lot of molecules parts of volume (1899). This phenomenon is called the **Rayleigh** (or **molecular**) **scattering** of light. Intensity of scattered beam in this case obeys the formula, very similar to (9):

$$I_{sc} = \frac{I_0}{\lambda^4} \cdot \left(1 + \cos^2 \theta\right) \cdot \frac{2\pi (n-1)^2 V}{N r^2} \qquad (10)$$

where V is the volume of air scattering light, N is the concentration of molecules. Formula (10) is also derived in the assumption of long light waves compared with fluctuated volumes of atmosphere. It is assumed in derivation of formulas (9) and (10) that if incident light is plane-polarized then scattered one is plane-polarized too.

A German physicist Gustav Mie (and a Dane Ludwig Lorenz independently) proposed a theory of light scattering for the case when the character size of impurities is of the order of light wavelength. They considered interference of incident light with light diffracted at the particles of impurities. **Mie-Lorenz** result (this mechanism is also called sometimes as **non-molecular scattering**) contains rather complicated formulas, which essentially depend on shapes of particles, therefore they are not presented here.

Both Rayleigh and Mie-Lorenz theories considered only **elastic** mechanisms of light scattering, when frequency (wavelength) of light does not change. **Inelastic scattering** was discovered by Raman and Krishnan from India (1928). Two Soviet scientists Mandelshtam and Landsberg declared that they made the same discovery in the same time independently, but they published their results later. Therefore this effect is called **Raman** (or **combinational**) **scattering** now and it consists in absorption of light of some frequency ω_0, which excites subsequent generation of EM waves with frequencies $\omega_0 \pm \Omega_i$, where Ω_i are inherent frequencies of molecular

oscillations. Because those inherent frequencies are character for every given substance, Raman scattering had become an important tool of investigation of a material structure.

Luis Brillouin considered also scattering of light on thermal oscillations of atoms or molecules of some substance (1922). Resultant frequency is expressed with a similar formula $\omega = \omega_0 \pm \Omega_t$, though this is an example of elastic scattering, the frequency shift in the case of **Brillouin scattering** appears as a result of light diffraction on acoustic waves (or, in other words, periodic changes of density of the substance). It became clear when Peter Debye had stated that thermal oscillations of particles in a substance could be considered as acoustic waves.

7. Light absorption

We know from the part 3 of the present Lecture that even substances, which are reputed to be transparent, can absorb light at their character frequency bands. (Light absorption does not violate energy conservation, because taken up energy is expended for heating of absorbing medium.)

Let us consider light propagation in an absorptive medium (*Fig. 7*). Suppose the incident intensity at $x = 0$ is I_0. At some depth *x* intensity is *I(x)*. We can also assume that amount of absorbed light energy in a very thin layer of the substance is proportional to the initial intensity at the input of this layer *I(x)* and to thickness of the layer *dx*:

$$dI = -\alpha \cdot I \cdot dx \qquad (11)$$

where α is the **absorption coefficient** of the medium, sign minus indicates that intensity of light diminishes in propagation through matter. (11) is an ordinary differential equation with separable variables. It can be easily solved, giving

$$I = -K \cdot e^{-\alpha x}$$

where *K* is an integration constant. It can be eliminated, using the above mentioned boundary condition, namely:

$$I\big|_{x=0} = I_0$$

So we obtain the **Bouguer law** of light absorption (1729):

$$I = I_0 \cdot e^{-\alpha x} \qquad (12)$$

Absorption coefficient of the substance is $\alpha = 4\pi\kappa/\lambda$, where κ is the attenuation coefficient (recall formula (8)), λ is the wavelength of light.

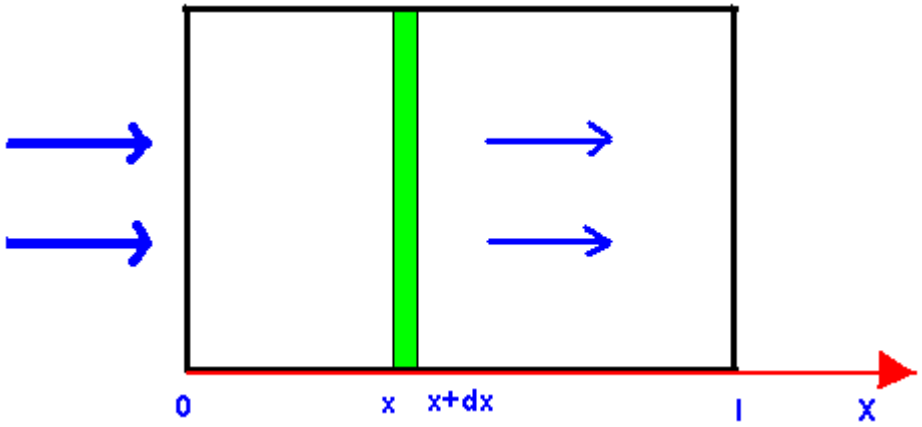


Fig. 7

Later Lambert and Beer reformulated Bouguer law (12) using the experimental fact that absorption coefficient of a solution is proportional to the solute's concentration *c*:

$$I = I_0 \cdot 10^{-\varepsilon c l} \qquad (13)$$

ε is called the molar absorption coefficient. (13) is the **Bouger-Lambert-Beer law**.

It should be mentioned that the absorption law in form (12) has more generality than (13). In fact, every process of loosing of energy in the process of propagation of any kind of waves at rather small intensities obeys (12). For example, change of energy of a light beam in the process of the Rayleigh scattering considered in the previous section can be described with equation (12) too.

LECTURE 10. THERMAL RADIATION. LUMINESCENCE.

1. Equilibrium radiation

Thermal radiation is a flux of electromagnetic waves, which are emitted by substances heated over absolute zero temperature. This kind of radiation has continuous spectrum and can exist in the state of thermal equilibrium with a surrounding matter.

Imagine that some heated subject is placed into a closed cavity, which does not pass electromagnetic radiation through its walls. This subject would radiate EM waves of different wavelengths. Many observations show that in some time multiple reflections of EM radiation from the walls of the cavity (associated, maybe, with absorption of part of light at every reflection) lead to establishment of equilibrium state of radiation in the cavity (*Fig. 1*). This **equilibrium radiation** would be uniformly distributed in volume of the cavity and directed with equal probability in all possible directions. Amount of radiated and absorbed energy would be equal for every small part of inner surface of the cavity as well as of the subject in the state of equilibrium. As it will be shown later, energy density of radiation would depend only on temperature.

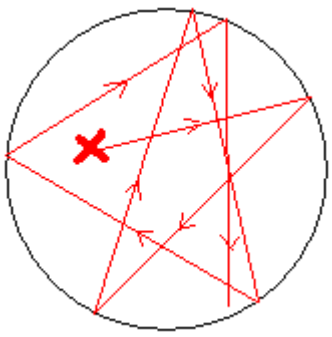


Fig. 1

One of the values characterizing equilibrium radiation is called its **spectral density**, that is amount of energy of radiation in unit volume related to an interval of wavelength (or cyclic frequency):

$$u_\lambda = \frac{\partial^2 W}{\partial V \partial \lambda}; \qquad u_\omega = \frac{\partial^2 W}{\partial V \partial \omega} \tag{1}$$

where W is the total EM energy in the cavity, V is the cavity volume. Spectral density depends on temperature in the cavity.

Two more notions are necessary for the further consideration. **Radiating capacity** (or **emissive power**) of some body is spectral density of a flux of radiant energy $d\Phi$ emitted in all directions by the body from unit area in unit time in a very narrow interval of wavelength $d\lambda$:

$$r_\lambda = \frac{d\Phi}{d\lambda} \tag{2}$$

Radiant emittance R_{EM} of a body is the total power of EM waves emitted from its unit surface in unit time. As thermal radiation has a continuous spectrum the radiant emittance is sum of values of emissive power at all wavelenghts

$$R_{EM} = \int r_\lambda \cdot d\lambda \tag{3}$$

Absorptive capacity (**absorptivity**) of a body is the part of energy of incident EM waves at a considered wavelength, which the body absorbs:

$$A = \frac{(\Phi)_{abs}}{(\Phi)_{total}} \tag{4}$$

Absorptive capacity cannot exceed unity according to definition. It is well known that absorptivity strongly depends on the wavelength, so usual window glass passes visible light but absorbs ultraviolet radiation almost completely. **Monochromatic absorptivity** characterizes absorptic properties in a narrow spectral interval and is equal to derivative of the latter value on wavelength:

$$a_\lambda = \frac{dA}{d\lambda} \tag{5}$$

2. Kirchhoff law

Consider a cavity like that drawn in *Fig. 1*. Suppose that the state of equilibrium is achieved. Than energy flux of EM radiation $\varepsilon_\lambda \cdot ds \cdot dt$ would fall at every small part of inner surface in cavity having area ds in time dt at some wavelength λ. As we consider the equilibrium state, the same flux is also emitted from the same area of cavity surface. Some portion of incident radiation of the wavelength λ would be absorbed by the considered part of the cavity, its amount is equal

$$a_\lambda \cdot W \cdot ds \cdot dt$$

Simultaneously, a heated wall of the cavity also radiates energy, equal to

$$R_{EM} \cdot ds \cdot dt$$

Let us equalize two fluxes of energy transmitted to the wall and from it:

$$r_\lambda \cdot ds \cdot dt + (1 - a_\lambda) \cdot \varepsilon_\lambda \cdot ds \cdot dt = \varepsilon_\lambda \cdot ds \cdot dt$$

So we obtain the equation called the **Kirchhoff law of radiation**:

$$\varepsilon_\lambda(\lambda, T) = \frac{r_\lambda}{a_\lambda} \qquad (6)$$

It states that ratio of radiating capacity to absorptive capacity is constant for all bodies and depends only on absolute temperature T and on wavelength λ (frequency).

Let us introduce here another one concept – **absolute black body** or **Planckian radiator**. This is an object, which has monochromatic absorptivity equal to unity for EM radiation of all wavlengths. Planckian radiator is some idealization, there are no real objects with such properties. For example, black soot is an excellent absorber of visible light, but infrared radiation penetrates through it. A good example of a Planckian radiator is a hollow cavity with a small window, through which light can propagate into the cavity and out of it (*Fig. 2*). Such cavities are used in experiments with absolutely black bodies.

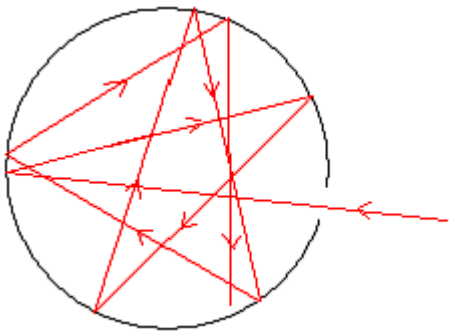


Fig. 2

It was found that in the state of thermodynamic equilibrium for absolute black body its radiant capacity is proportional to the spectral density of thermal radiation:

$$\varepsilon_\lambda(\lambda, T) = \frac{c}{4} u_\lambda \qquad (7)$$

where c is the speed of light (EM wave) in vacuum. Or, considering dependence on frequency, but not on wavelength we have

$$\varepsilon_\omega(\omega, T) = \frac{c}{4} u_\omega \qquad (7, \text{a})$$

As absorptive capacity of a Planckian radiator is equal to 1, it means that value ε_λ is emissive power of an absolutely black body. There are some practical subsequencies for the Kirchhoff law:

1) *Planckian radiator has maximal radiating capacity at a given temperature.*

2) *If some object weakly absorbes electromagnetic radiation in some frequency range (for example, reflects visible light), than it radiates in the same also very weakly*.

It should be also mention the objects called **gray bodies** – they have constant absorptivity in some rather wide interval of frequencies (wavelengths). For example, human skin can be considered as a gray body for infrared waves ($a_\lambda \approx 0.9$).

3. Stefan-Boltzmann law. Wien displacement law.

Experiments on determination of radian capacity of a Planckian radiator had carried out by John Tyndall and Josef Stefan. The latter generalized experimental results and found (1879) that radiant emittance of absolute black body depends on temperature as

$$\left(R_{EM}\right)_{bb} = \int \varepsilon_\lambda \cdot d\lambda = \sigma T^4 \qquad (8)$$

Five years later Ludwig Boltzmann had confirmed this result theoretically, therefore the constant $\sigma = 5.66 \cdot 10^{-8}\ \ Wt/\left(m^2 \cdot K^4\right)$ is named after both Stefan and Boltzmann. For gray bodies in the range of constant absorptivity that constant is replaced with the **reduced coefficient of absorption** $\delta = \alpha \cdot \sigma$, so that

$$\left(R_{EM}\right)_{gray} = \delta T^4, \tag{9}$$

It should be specially emphasized that the Stefan-Boltzmann law (8) is valid only for absolutely black bodies. In an object is neither absolutely black nor gray body then coefficient of proportionality between radiant emittance and the forth power of absolute temperature often depends on the wavelength of radiation, so a formula similar to (8) cannot be written at all.

Wilhelm Wien had proved that dependence of emissive power on the wavelength of emission has maximum for a black body at the next wavelength (1893):

$$\lambda_{\max} = \frac{b}{T} \tag{10}$$

where **Wien constant** is equal to $b = 2.89 \cdot 10^{-3}\ m \cdot K$.

The Wien's law (10) can be correctly applied only to Plackian radiators.

4. Planck formula

When analyzing experimental data on radiation and absorption of EM waves Wilhelm Wien stated (1894) that spectral density of thermal radiation in equilibrium can be expressed as

$$u_\omega(\omega, T) = \omega^3 F\left(\frac{\omega}{T}\right) \tag{11}$$

Some years later (1905) James Hopwood Jeans derived the expression for the spectral density of equilibrium thermal radiation:

$$u_\omega = \frac{kT}{\pi^2 c^3}\omega^2 \tag{12}$$

Formula (12) is called the **Rayleigh-Jeans law**, because Jeans used in his derivation Rayleigh's idea to consider electromagnetic radiation in a closed cavity as a superposition of infinite quantity of standing waves with different wavelengths (and frequencies), so that mean energy of any standing wave is equal kT - in accordance with classical thermodynamics. If (12) were absolutely correct, it would give the formula of the function $F\left(\omega/T\right)$ from (11).

But experiments showed that Rayleigh-Jeans law is valid only for long wavelengths. Moreover, attempts to calculate radiant emittance of an absolute black body using (12) and (7) and carrying out integration on frequency from zero to infinity leads to $(R_{EM})_{bb} \to \infty$, but not to the Stefan-Boltzmann dependence (8).

Max Planck (1900) proposed to save the day when he had assumed that electromagnetic waves are radiated not continuously, but in small portions, called him guanta (plural of **quantum**). Light quanta Einstein later called **photons**. Energy of a photon should be equal by his hypothesis to:

$$E = h \cdot f \tag{13}$$

where $h = 6.626 \cdot 10^{-34}\ J \cdot s$ is called now the Planck constant, $f = \omega / 2\pi$ is the linear frequency of a wave. Nowadays formula (13) is often used in a bit different form with $\eta = h / 2\pi$:

$$E = \eta \cdot \omega \tag{13, a}$$

Though Planck solved this problem in 1900, we consider Einstein's derivation (1916) of Planck's formula for spectral density of thermal radiation, because it will be useful in our future study of operation of lasers. Einstein used essentially Bohr's postulates formulated in 1913, which we mentioned in the previous Lecture. Following Bohr and Einstein we know that electromagnetic wave can be generated as atom passes from some higher energy level E_m to a lower energy level E_n and energy of emitted photon is equal in this case to (this is the **Bohr frequency rule**):

$$E = \eta \cdot \omega = E_m - E_n \tag{14}$$

But the photon's energy is used to change a lower atomic energy level E_n to a higher one E_m in the opposite case of absorption of EM wave. Suppose EM radiation passes through a substance containing many atoms with energy levels E_n and E_m, suppose also that N_n and N_m is population of those levels. Then it is natural to assume that amount of transitions of atoms from the higher level downward should be proportional to the number of atoms at that higher level: $A_m^n N_m$. Einstein has named EM waves emitted in such transitions the **spontaneous radiation**. Number of transition in the opposite direction should be proportional to the number of atoms at the lower level as well as to the spectral density of radiation at the

frequency corresponding to that transition: $B_n^m N_m u_\omega(\omega_n^m)$, where $\omega_n^m = (E_m - E_n)/\eta$. The main significance of that Einstein's work is in hypothesis of existence of **induced radiation** of EM waves: if atom, being initially at the energy level E_m, absorbs a photon with energy $\eta\omega_n^m$ then, according to Einstein's assumption, the transition to lower energy level also takes place, but as a result two absolutely identical photons $\eta\omega_n^m$ are radiated. Number of induced transitions in the considered system of many atoms is $B_m^n N_m u_\omega\left(\omega_n^m\right)$. Parameters A_m^n, B_n^m, B_m^n are called **Einstein's coefficients**, they depend on the nature of an atom and on the frequency of the transition ω_n^m. In the equilibrium state

$$A_m^n N_m + B_m^n N_m u_\omega\left(\omega_n^m\right) = B_n^m N_n u_\omega\left(\omega_n^m\right) \tag{15}$$

It can be proven that $B_m^n = B_n^m$, therefore

$$u_\omega\left(\omega_n^m\right) = \frac{A_m^n N_m}{B_m^n (N_n - N_m)} = \frac{A_m^n}{B_m^n\left(N_n / N_m - 1\right)}$$

If electrons in atoms of the considered substance obey the Boltzmann distribution, then

$$\frac{N_n}{N_m} = \exp\left(\frac{E_m - E_n}{kT}\right)$$

where *k* is the Boltzmann constant. Combining two latter formulas we obtain

$$u_\omega\left(\omega_n^m\right) = \frac{A_m^n}{B_m^n\left[\exp\left(E_m - E_n / kT\right) - 1\right]} \tag{16}$$

Comparing Eq. (16) with (11) we can confirm Planck's preposition formulated in form of the Bohr's postulate (14):

$$u_{\omega}\left(\omega_n^m\right)=\frac{A_m^n}{B_m^n\left[\exp\left(\eta\omega_n^m/kT\right)-1\right]} \qquad (16, a)$$

Expanding exponent from (16, a) in series and retaining only linear term, we have from the latter the next formula

$$u_{\omega}\left(\omega_n^m\right)=\frac{A_m^n kT}{B_m^n \eta\omega_n^m} \qquad (16, b)$$

For long wavelengths (or small frequencies) formula (16, b) should coincide with the Rayleigh-Jeans law (12), which surely complies experimental results under these conditions. So comparing those results we can get for Einstein coefficients

$$\frac{A_m^n}{B_m^n}=\frac{\eta\left(\omega_n^m\right)^3}{\pi^2 c^3} \qquad (17)$$

This expressioncan can be interpreted as ratio of probability of spontaneous radiation to that of induced one. Substituting formula (17) in (16, a) we obtain:

$$u_{\omega}\left(\omega_n^m\right)=\frac{\eta\left(\omega_m^n\right)^3}{\pi^2 c^3\left[\exp\left(\eta\omega_n^m/kT\right)-1\right]}$$

Real atomic system contains many energy levels and electromagnetic radiation can have wide spectrum, therefore indices *m* and *n* can be omitted in this expression and we obtain the **Planck formula** for spectral density of thermal radiation in equilibrium:

$$u_{\omega}(\omega)=\frac{\eta(\omega)^3}{\pi^2 c^3\left[\exp\left(\eta\omega/kT\right)-1\right]} \qquad (18)$$

Substitution (18) in formula (7) and integration thus obtained expression on cyclic frequency from zero to infinity, leads to

$$\left(R_{EM}\right)_{bb} = \frac{\pi^2 k^4}{60 c^2 \eta^3} T^4 \qquad (19)$$

Formula (19) expresses the Stefan-Boltzmann constant through the "more fundamental" parameters π, c, k, η.

A problem. Try to derive Wien's constant from (18).

5. Luminescence

Luminescence is a glow that cannot be explained by thermal radiation. Therefore its wavelength does not depend on temperature and its intensity in many cases depends on chemical composition. Sometimes luminescence is also called a cold glow. Electromagnetic radiation is generated in some very short time interval after a special substance called **phosphor** (or **luminophor**) had been exposed to an external action.

Depending on the mechanism of excitation the next kinds of luminescence are distinguished:

photoluminescence. Active substance irradiates light after exposure with visible or ultraviolet light;

roentgenoluminescence. The same as previous, but initial action carried out with X-rays;

cathodoluminescence (inverse photoelectric effect). Phosphor radiates EM waves after bombardment with electron beam;

radioluminescence is observed under influence of products of radioactive decay, such as $\alpha-$particles, $\beta-$particles, $\gamma-$rays, protons and so on;

chemiluminescence is light emission at some chemical reactions (photo in *Fig. 3* from the Wikipedia);

Fig. 3

electroluminescence is called excitation of gas molecules by electric discharge;

bioluminescence is glowing of breathers (more than 200 species) as a result of their vital activity;

triboluminescense – glowing at destruction of some crystals (sugar, for example);

sonoluminescence is light emission during propagation of high-intensity sound waves in some solutions

and some others.

The mechanism of the photoluminescence is understandable with use of one of Bohr's postulates mentioned above. A photon of external electromagnetic radiation is absorbed in a substance. That leads to transition of one electron from a lower to a higher energy level. Three different cases are possible further:

a) The atom undergoes collisions with its neighbors and loses part of obtained energy on change of temperature of the substance. The other part of energy is eliminated in form of electromagnetic wave with wavelength longer than that of excited wave. This is a kind of **Stokes luminescence**, named after George Stokes, who formulated so-called Stokes rule, valid for substances observed in his experiments: phosphorus eliminates light with wavelength longer than that of absorbed light: $\lambda_{lum} > \lambda_{abs}$. In fact, this is a result of conservation of energy.

b) **Anti-Stokes luminescence** ($\lambda_{lum} < \lambda_{abs}$) takes place in other substances. Energy conservation law does not violate there, if an incident photon is absorbed by atom being at an excited energy level and glowing takes place in the following transition to the ground level.

c)In some cases atom emits a photon with the same wavelength as the absorbed photon had. This is the case of **resonant glowing**, which was experimentally considered by Robert Wood (1904-1905).

Study of spectrum and intensity of luminescence or **fluorimetric analysis** is used in chemistry and biology for detection of presence (or absence) of definite substances in samples under consideration. Comparison of intensity of luminescence with the intensity of a reference sample allows chemists to determine quantity of the searched substance with accuracy up to 10^{-10} *gram*.

Addition of some phosphorescent molecules to a substance under investigation (**luminescent marks**) helps to trace motion of molecules in chemical processes.

LECTURE 11. LASERS AND THEIR APPLICATIONS

1. Induced radiation

In the previous Lecture we mentioned Einstein's equation for an equilibrium process of absorption and generation of light:

$$A_m^n N_m + B_m^n N_m u_\omega\left(\omega_n^m\right) = B_n^m N_n u_\omega\left(\omega_n^m\right) \qquad (1)$$

We said that the first term in the left-hand side of Eq. (1) describes **spontaneous** transitions from a higher energy level E_m to a lower one E_n accompanied by radiation of a photon with energy $E_{mn} = \eta \cdot \omega = E_m - E_n$. The second term in (1) corresponds to **induced radiation:** as atom, being at the excited energy level E_m, absorbs a photon of external EM radiation with energy E_{mn}, by Einstein's hypothesis (later confirmed by construction of lasers), a couple of two photons can be emitted and those photons are absolutely identical to the incident one. They have the same frequency, direction of propagation and polarization as the initial photon-trigger, their length of coherence is much longer than of spontaneously emitted electromagnetic waves. Maybe it was a Soviet physicist Valentin Fabrikant who first proposed to use induced emission to produce coherent, intense and monochromatic light (1939). Alexander Prokhorov, Nikolay Basov (both USSR) and Charles Townes (USA) wined the 1964-year Nobel prize on physics after there independent preposition of design of sources of coherent optical waves.

Some evident difficulties were seen in this way. First of all, as we also mentioned in the Lecture 10, relation of probabilities of spontaneous and of induced radiation is proportional to the cube of frequency:

$$\frac{A_m^n}{B_m^n} = \frac{\eta\left(\omega_n^m\right)^3}{\pi^2 c^3} \qquad (2)$$

Numerical estimations show that in the radio-wave frequency range ratio (2) is about 10^{19} times less than in optical range. Therefore the first devices for generation of induced electromagnetic radiation emitted centimeter waves (USA, 1954). They were called **masers** (**m**icrowave **a**mplification by **s**timulated **e**mission of **r**adiation) by one of principal their designers Charles Townes. Later astronomers have found that giant gas clouds in cosmos can emit coherent induced radiation, there were in fact

natural masers. Masers are not widely known in our days, though they are used as very precise frequency standards.

There is also another one difficulty in practical implementation of stimulated radiation of optical wavelength. To achieve amplification of an induced wave it is necessary that population of a higher energy level N_m would exceed population of a lower level N_n. In equilibrium state it is impossible: distribution of electrons in atoms obeys the Boltzmann law:

$$N_i \sim \exp\left(-\frac{E_i}{kT}\right)$$

It is possible to reach in some medium the non-equilibrium thermodynamic state with respect to energy levels E_m and E_n, when $N_m > N_n$ (this state is called **inverse pop ulation** of energy levels). A medium, in which inverse population is possible, is called **active**. One external photon with energy $E_{mn} = E_m - E_n$ in an active substance is enough to reach many its twins. Active medium in a laser (***l***ight ***a***mplification by ***s***timulated ***e***mission of ***r***adiation) is usually placed between two parallel mirrors to cause emitted photons travel several times back and forth to embrace more atoms from the excited state E_m in generation. As intensity of generated light reaches some threshold value a special optical shutter releases generated light beyond the laser. First operating laser was made by Theodore Maiman in the USA using a ruby crystal as active medium (his results were published in August of 1960).

2. Ruby laser

Ruby is a crystalline aluminium oxide (Al_2O_3), in which less than 1% of aluminium is replaced by chromium. Ruby crystals can be grown artificially with lengths up to some dozens of centimeters and this is an advantage of that material as the active medium. Besides that ruby is a birefringent crystal. If its optical axis does not coincide with the longitudinal axis of a laser, generated light is (partly or completely) plane-polarized.

Energy levels of chromium are schematically shown in *Fig. 1*.

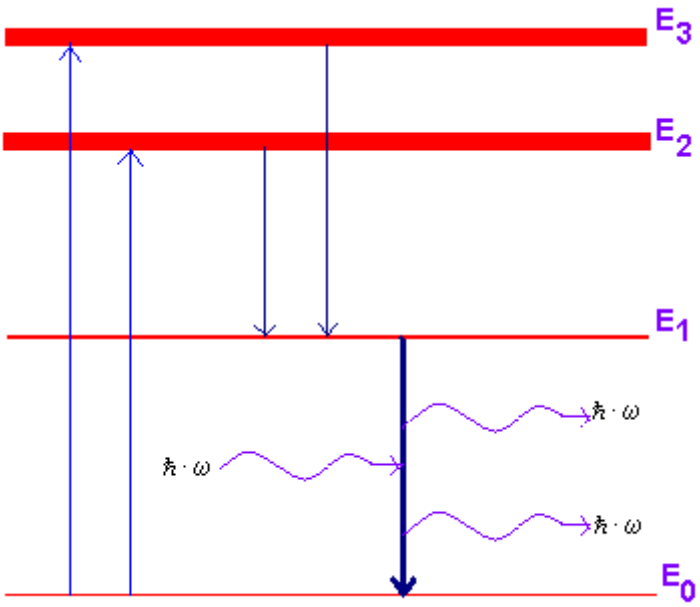


Fig. 1

The next generally accepted convention is used when depicting *Fig. 1*: the shorter is lifetime of an atom at a certain energy level the wider line indicates it. (In fact, this is graphical demonstration of the uncertainty principle, which will be considered in the Lecture 13). Energy level E_0 is stable and in the state of thermodynamic equilibrium absolute majority of atoms in ruby have this value of energy. Levels E_2 and E_3 are unstable (lifetime is of the order of 10^{-8} *second*), while the level E_1 is **metastable** (an atom can linger at this level in ruby for about $2 \cdot 10^{-3}$ *second*).

The **optical pumping** is used to create inverse population between levels E_1 and E_0, that is a short flash of a powerful white lamp eliminates a ruby crystal. It should be mentioned that creation of inverse population in a two-level medium is impossible. Indeed, probability of transitions B_m^n and B_n^m are equal and as soon as N_1 matches N_0 the rate of induced radiation becomes the same as the rate of pumping. Therefore optical pumping in ruby induces transitions from the ground level E_0 to the unstable levels E_2 and E_3. Their relative breadth increases efficiency of pumping. It was calculated theoretically that probability of inverse transitions $E_3 \to E_0$ and $E_2 \to E_0$ is negligible compared with probabilities of radiationless transitions from those unstable orbits to the metastable level E_1. (Energy released in those transitions is dissipated in form of heat.) Relatively long lifetime at this level allows to accumulate there more than half of cromium

atoms. Energy of those atoms is emitted in form of a red light pulse ($\lambda = 0.6943\ \mu m$) in the radiative transition $E_1 \to E_0$ at the final step.

Two opposite sides of a ruby crystal are sometimes polished, so they are used as mirrors. Ruby lasers are very powerful, values of their output reach up to $10^7 Watt$. But impossibility of continuous operation is their disadvantage.

3. Other types of lasers

Besides the three-level active media (like ruby crystals), the four-level materials are also used (such as CaF_2 with some admixture of rare-earth atoms). In such case inverse population is usually created between two different metastable levels. That allows to use a not very intense pumping and operate in continuous mode.

Very important, widely used type of laser is the He-Ne one, which was first demonstrated by Javan, Bennet and Erriott in December 1960. These gases are placed in a discharge tube at very low pressure (of the order of 1 *mm Hg*, partial pressure of helium is 9-10 times more than that of neon). Inverse population is created between a metastable level of He and one of unstable levels of Ne. Pumping is implemented by applying high voltage between electrodes in the tube. Collected at the metastable level helium atoms excite very close (in energy scale) unstable levels of neon in their thermal motion. Red-light emission ($\lambda = 0.6328\ \mu m$) takes place as neon atoms pass from there to another one unstable energy level.

He-Ne lasers are able to operate in continuous mode, they have excellent monochromaticity and radiation directivity. Linear polarization of light is attained with the help of galss plates closing both sides of a gas tube and maintained at the Brewster angle to light propagation direction.

Dozens of different types of lasers are designed in last decades, they have different color, size, power, but active media for the first types of lasers were chosen so successfully, that ruby and He-Ne lasers have their niche in applications until now.

4. Properties of laser light

1. The main property of laser light is its very high coherence. It has the next explanation: natural is generated in spontaneous transitions from unstable energy levels with lifetime $\tau \sim 10^{-8} s$), therefore length of one portion of emitted natural light (coherence length) is of the order of

$$l_{coh} = c \cdot \tau \sim 3 \ m$$

Lifetime of an atom at a metastable level is much more than dozens of nanoseconds, therefore coherence lengths of some modern lasers can reach thousands of kilometers.

2. Laser light is highly collimated because light beam usually passes between mirrors several times until its escape from active medium. That is why broadening of a laser beam is very weak at long distances of propagation. Loss of intensity is correspondingly small too.

3. Laser light is very monochromatic becausel of generation from metastable levels and also because of interference between mirrors.

4. Laser light is plane-polarised.

5. Laser light is very intensive compared with natural one. So intensity of some lasers reaches values up to $10^{16} \ Wt/m^2$, while the sunlight intensity at the Earth surface does not exceed 0.1 Wt/m^2.

6. Because of its high intensity and collimation laser light exerts high pressure. It is known that pressure of a light beam of intensity I, which is incident normally at a completely absorbing surface, is $p = I/c$ (c is the light velocity). So we obtain:

$$p_{laser} = \frac{10^{16}}{3 \cdot 10^8} \approx 3 \cdot 10^7 \ Pa \quad \text{versus} \quad p_{natural} = \frac{10^{-1}}{3 \cdot 10^8} \approx 3 \cdot 10^{-10} \ Pa$$

It should be emphasized that such properties of laser light are the results of its principally different physical origin compared with natural radiation.

5. Holography

Idea of holography as a way of making of three-dimensional images had been first proposed and experimentally checked by a Polish physicist Mieczyslaw Wolfke in 1920, but his work had anticipated time and was forgotten soon. Repeatedly the holography principle was suggested by an English engineer Dennis Gabor in 1947, but its practical implementation was impossible until appearance of lasers. First high-quality holograms of objects with awkward shape were demonstrated by Emmeth Leihth ans Juris Upatnieks in 1960-1963.

The simplest technology of holography can be described in the next way. Process of hologram recording is shown in *Fig. 2*.

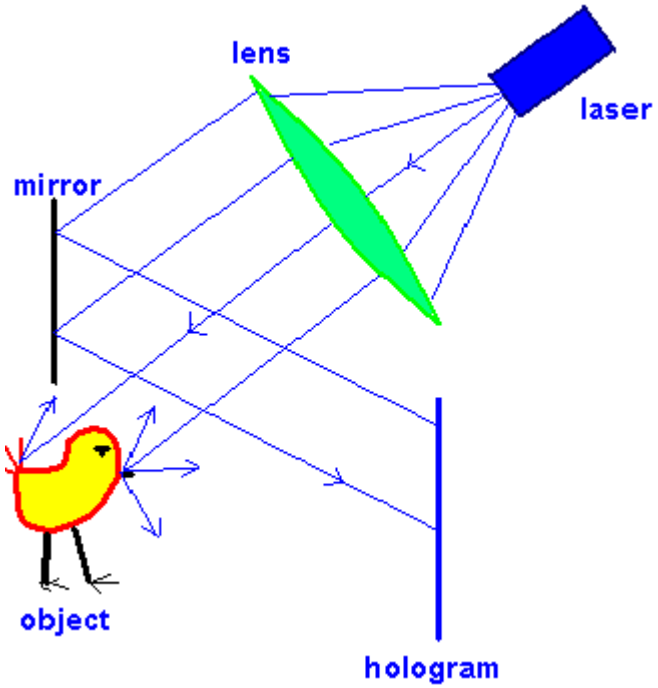


Fig. 2

Laser beam is broaden with the help of lens, half of it ("object beam") is scattered at the object and illuminates a photoplate, forming a pattern of different intensity (at a first glance like in the process of usual photography). But our 3D perception of objects is based on phase measurements too, as our brain estimates difference of distances between an object and every of two eyes and this helps it to graw the 3D image. In holography method phase difference between rays arrived at the photoplate from different points of the object is made by means of another one ("reference") beam, which is usually produced with the same laser, reflects from a mirror and interferes with the object beam at the holographic photoplate.

A principle scheme explaining reading of a hologram is drawn at *Fig. 3*.

In this case the hologram is placed where mirror was staying in recording process and the holographic photoplate is illiminated with the same laser. An observer can look at the virtual 3D image of initial object through the hologram plate. Real image is usually not used, because it is mirror-inverted with respect to object. This virtual image is optical illusion: it can be photographed at different angles and the object at snapshots would be seen from different viewpoints.

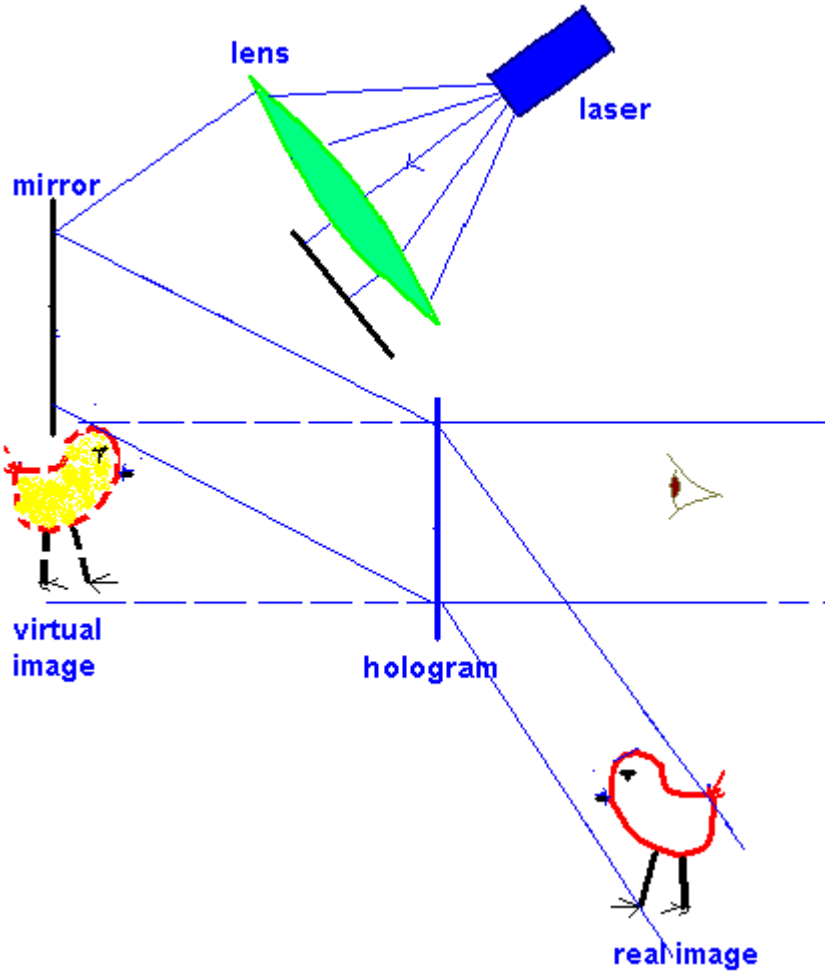


Fig. 3

It should be specially emphasized that coherence length between the object beam and the reference one in the process of a hologram recording can be up to dozens meters, therefore practically useful holograms were impossible until appearance of industrial lasers.

Holography is a widely used technology with numerous applications. Everybody has held holograms in own hands, when using banknotes or plastic cards, which are defended with the help of holographic images. Medical application of holography are mentioned in the next paragraph.

6. Applications of lasers in chemistry and medicine

Mentioned above properties of induced radiation cause their numerous applications. All processes using lasers in chemistry, biology and medicine, in fact, can be divided into three groups:

1) nonperturbative effects, which use coherence or monochromatism of induced radiation;

2) photochemical effects based on laser excitation of reagents;

3) thermal action at reaction products or biological tissues.

High monochromatism of laser light is used in chemistry to influence only at some selected atoms with definite energy levels to influence on the progress of chemical reactions. Sharp directivity of a laser beam allows to effect at a very small volume of reagents.

Medical applications of laser light embrace use in diagnostics, therapy and surgery.

Interference of coherent light waves brings information about surfaces of biological objects (**speckle interferometry**).

Three-dimensional images of viscera are obtained by means of **holography**, which essentially uses high coherence of laser light.

Registration of angular dependence of intensity of scattered light from an incident narrow beam allows estimate sizes of small particles (**nephelometry**).

Measurement of frequency shift of a monochromatic beam incident at moving particles (bacteria, blood cells) helps to determine velocity of their movement (**Doppler laser anemometry**).

Intense laser light can evaporate matter from the surface of a tissue. Spectral analysis of the resultant fume provides information about chemical composition of substance (**mass spectrometry**).

He-Ne-laser radiation is used for medical treatment of wounds, ulcers because red light is believed to help healing.

Blue laser light is used for cure of icterus of newborn children. This disease is related with high concentration of bilirubin in organism. Bilirubin strongly absorbes blue light and dissociates under its exposure.

Induced radiation is used in surgery for dissection of tissues, removal of pathological areas, stopping of bleeding, welding of biotissues.

Advantages of laser surgery are:

1) contactlessness and therefore sterility;

2) selectivity based on monochromatism (maximal absorbtion of neighboring tissues takes place at different wavelengths);

3) exsanguinity (because of thermal coagulation of proteins);

4) possible microsurgery (due to sharp focusing of a beam).

7. Laser tweezers

In 1970 Arthur Ashlee from the USA published results of his experiments on optical trapping of small particles (in 2019 he have been

awarded with the Nobel prize for this idea and its realization). Explanation of principle of catching of a particle in a focused optical beam near its waist is seen in *Fig. 4*. (The part of a beam with the smallest diameter is called the waist.)

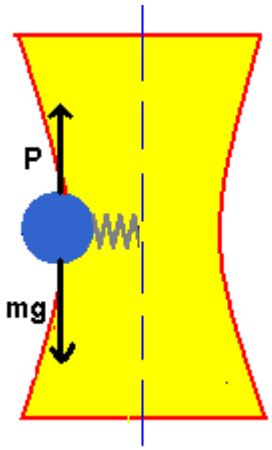


Fig. 4

Light is propagated upwards at *Fig. 4*, and the gravitational force and the light pressure act at the dielectric particle caught with the beam (the particle is denoted as the ball at the figure). If these forces are opposite to each other (like it is shown in *Fig. 4*) then a phenomenon of optical levitation of the particle is observed.

Electric field in the waist of the beam is very strong and non-uniform, therefore if the trapped particle changes its position from the centeral vertical axis of the beam then a restoring force acts at it like a string in an experiment demonstrating the Hook's law (Lecture 2). This force is of electrostatic origin (electric dipoles are taken in the direction of electric field, Lecture 5).

Different applications of laser tweezers are proposed in the last decades. They are used in works on quantum computing and biophysics, where separate cells are processed using this method. Artificial composition of a single molecule from separate atoms trapped by laser beams is proposed in some recent publications (2020).

Further development of this area promises many interesting results.

LECTURE 12. X-RAY RADIATION

1. Sources of x-radiation

X-ray emission had been discovered by Wilheln Konrad Röntgen in Germany in 1895, therefore sometimes it is called röntgen radiation too. He had found that a gas-discharge tube covered with cardboard to prevent propagation of visible and ultraviolet light nevertheless emits some radiation causing glowing of a luminescent material. In 1905 Charles Barkla proved experimentally that if this radiation consisted of waves they should be transversal. In 1912 Max von Laue with his co-workers observed diffraction of x-radiation on crystals. Now we know that x-rays are ordinary electromagnetic waves with wavelengths from 100 to 10^{-3} *nm*. At the scale of electromagnetic waves they are ranked between ultraviolet waves and γ – radiation. X-rays are studied separately because due to their wavelengths they have some distinctive properties and find important applications different from UV-waves or γ – rays.

X-radiation is emitted in radioactive transitions of some unstable isotopes of chemical elements like ^{55}Fe, for example, these are natural sources. X-ray tubes are used as artificial sources. A scheme of a usual X-ray tube is drawn in *Fig. 1*.

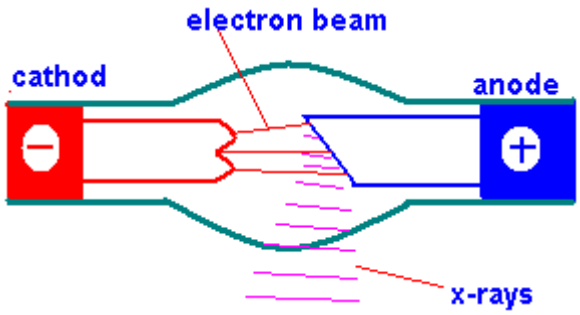


Fig. 1

The tube is an evacuated bulb of some dozens centimeters length containing two electrodes with high voltage (1 – 500 *kV*) between them. Cathode is heated with the strong electric current. Heated metal of the cathode emits electrons, which are accelerated by the applied voltage and bombard anode with very high kinetic energy. In this process two different kinds of X-rays are radiated, they are the **slowing-down radiation** (**bremsstrahlung** as it is often called following traditional German name) and the **characteristic radiation**. Electron bombardment of the anode heats it very intensively, therefore positive electrodes are usually made of (or they

include a plate made of) high-melting material like tungsten. The working surface of the anode is usually inclined with respect to the longitudinal axis of the tube to direct X-ray beam out of the tube.

2. Bremsstrahlung (slowing-down x-radiation)

Origin of the Bremsstrahlung can be easily explained with Maxwell's classical electrodynamics. As electrons when striking anode rapidly loose their high speed, they should generate electromagnetic waves. Slowing-down X-radiation has the next properties:

1) As any other EM wave X-rays are emitted by separate quanta. Every quantum has energy equal to

$$E = h \cdot f = \frac{hc}{\lambda}$$

where h is the Planck constant, as usual, c is the speed of light in vacuum, f and λ is the frequency of radiation and its wavelength respectively.

2) All electrons arriving to an anode have equal kinetic energy because they are accelerated from zero speed by the same electric field:

$$\frac{mv^2}{2} = eV$$

where m and e are electron's mass and charge, v is its velocity near the anode, V is voltage in a tube.

3) Kinetic energy of an electron is randomly divided into two parts: heating of the anode Q and energy of a radiated photon:

$$\frac{mv^2}{2} = Q + \frac{hc}{\lambda} \qquad (1)$$

Thus the Bremsstrahlung has a continuous spectrum (*Fig. 2*) , as different electrons emit photons of different wavelengths. And it is evident from Eq. (1) that wavelengths shorter than

$$\lambda_{\min} = \frac{hc}{eV} \qquad (2)$$

cannot be generated at a given voltage V. In practice this formula is used in the next catchy form:

$$\lambda_{\min} = \frac{1.24}{V} \qquad (3)$$

It is implied that voltage in (3) is measured in *kilovolts*, while wavelength in *nanometers*. This value does not depend on the material of an anode.

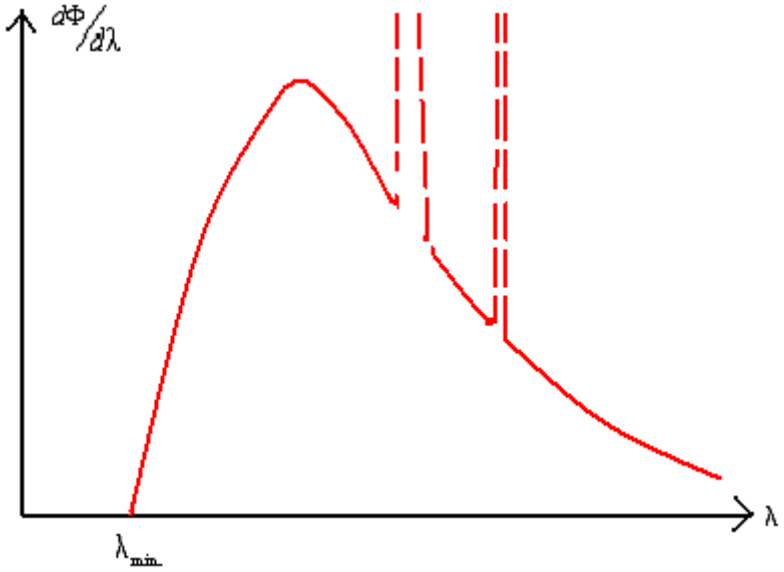


Fig. 2

If Φ is the flux of slowing-down X-radiation obtained with some tube, then dependence of its spectral density $d\Phi/d\lambda$ on wavelength typically lookes like a continuous curve in *Fig. 2*. It was also found that the flux depends on applied voltage, intensity of electric current in a tube and on the atomic number of a chemical element, from which the anode is made:

$$\Phi = k \cdot I \cdot V^2 \cdot Z$$

where coefficient $k \sim 10^{-9}\ Volt^{-1}$.

It should be also mentioned that the flux of X-ray emission from a tube grows as temperature of the cathode increases.

3. Characteristic x-radiation

Charles Barcla first caught sight of discrete narrow spectral lines, which were often seen at the background of the continuous spectrum. They narrow peaks are schematically drawn with dashed lines in *Fig. 2*. Barcla also found that wavelengths of those narrow peaks are different for different

materials used as a positive electrode. It was similar to optical spectra of gases therefore this phenomenon was called **characteristic X-ray spectrum**.

Characteristic X-radiation cannot be excited in bombardment of anode by protons or by $\alpha-$particles, it is also needs sufficiently high accelerating voltage, because there is some threshold of its radiation, depending of emitting substance. Characteristic spectrum is explained as a result of liberation of an electron from some inner shell (*K*-shell, for example) in atom in the process of the anode bombardment in the tube. Vacancy at this shell is quickly filled with another electron from a higher energy *L*-level, quantum of X-radiation is emitted simultaneously. Thus arose vacancy at the *L*-level is filled with another electron from yet higher shell and so on. So a set of lines is triggered with hit of only one electron.

Henry Moseley considered characteristic spectra of 38 chemical elements and stated that square root of frequency of a *K*-shell line is proportional to the atomic number of the appropriate element (1913):

$$\sqrt{f} = M \cdot (Z - a) \qquad (4)$$

where *M* and *a* are constants. He had also found that similar formulas, but with different values of *M* and *a*, are valid for *L*-lines, *M*-lines and so on. Equation (4) is called the **Moseley law**.

Moseley's experiments showed that the main parameter, which defines a position of an element in the periodic table, is its atomic number, but not the atomic mass (discerning for different isotopes). He arranged more precisely the order of some elements in the table. For example, in the beginning of the XX century it was not seen how to order cobalt and nickel.

It is interesting that there are no characteristic absorption spectra in the X-rays range. So if a beam of the slowing-down radiation passes through any substance no dark lines would be at character wavelength. It takes place because the Bremsstrahlung does not excite atomic energy levels and vacancies at the inner energy levels are filled.

4. Absorption of x-radiation

Interaction of X-radiation with a substance carries in one of the next three mechanisms.

1) If energy of an incident photon is rather small than **elastic scattering** takes place: energy (and the wavelength) of the photons remains the same, but direction of propagation can be changed. (Rayleigh scattering of light is an example of elastic scattering of optical photons.)

2-3) At higher values of the photon's energy it is possible **photoelectric effect** or **Compton scattering**. In the case of photoelectric effect energy of incident photon is completely expended for electron liberation from an atom. In the case of the Compton effect (or inelastic scattering) a part of photon's energy is used also for liberation of an electron, but a photon with smaller energy (longer wavelength) is radiated.

Absorptivity of X-rays in different substances is irrelevant to absorptivity of visuable light in the same materials. Lead glass (transparent in optical range) with some millimeters of thickness practically completely absorbes X-radiation, while a thin aluminium foil does not impeed it. The Bouguer law from the Lecture 9 is valid for X-rays too, so the decreasing flux of propagating through a substance X-wave obeys the formula:

$$\Phi(x) = \Phi_0 \cdot \exp(-\mu x)$$

where μ is called the **linear attenuation factor**. It was found that this value can be separated on different mechanisms of attenuation of X-radiation:

$$\mu = \mu_e + \mu_{ph} + \mu_c$$

where μ_e, μ_{ph}, μ_c are elastic, photoelectric effect and Compton attenuation factors respectively.

The linear attenuation factor is proportional to the mass density of a substance, therefore the **mass coefficient of attenuation** is also used:

$$\mu_m = \frac{\mu}{\rho} \qquad (5)$$

The next empiric expression turned to be sufficiently exact:

$$\mu_m = CZ^4\lambda^3 \qquad (6)$$

Eq. (6) explains formation of bones' images in medical examination of inner organs with the help of X-rays (fluorography). When it is necessary to discern some soft tissue among others a contrast substances (like $BaSO_4$) is gulped down by a fluorography patient.

5. Some applications of x-radiation

Numerous applications of X-rays are based on their ability to pass through substances opaque for visible light. First usage of newly discovered

radiation Röntgen demonstrated in the last decade of XIX century when he had made X-rays snapshot of his wife's hand. Broad application of this method of diagnostics helped army doctors in the first world war.

Nowadays this became an important tool of prevention of terrorist attacks in crowds.

Medical applications of laser light embrace use in diagnostics, therapy and surgery.

6. X-rays structure analysis

Max von Laue's idea of observation of diffraction of X-radiation on crystals turned out to be very fruitful. Average distances between atoms in solids are of the order of some angstroms and range of wavelengths of X-rays includes these values, it means that periodic structures of atoms form natural diffraction grating for this range of electromagnetic waves. If the angle between initial direction of propagation of the monochromatic wave and the layer of the atoms in the crystalline lattice is Θ then the condition of a maximum in the resultant diffraction pattern obeys **the Bragg equation**:

$$2d \cdot \sin\Theta = n\lambda \qquad (7)$$

where d is the distance between layers of atoms in the crystal lattice, n is a whole number, λ is the wavelength of X-rays (*Fig. 3*, from the Wikipedia).

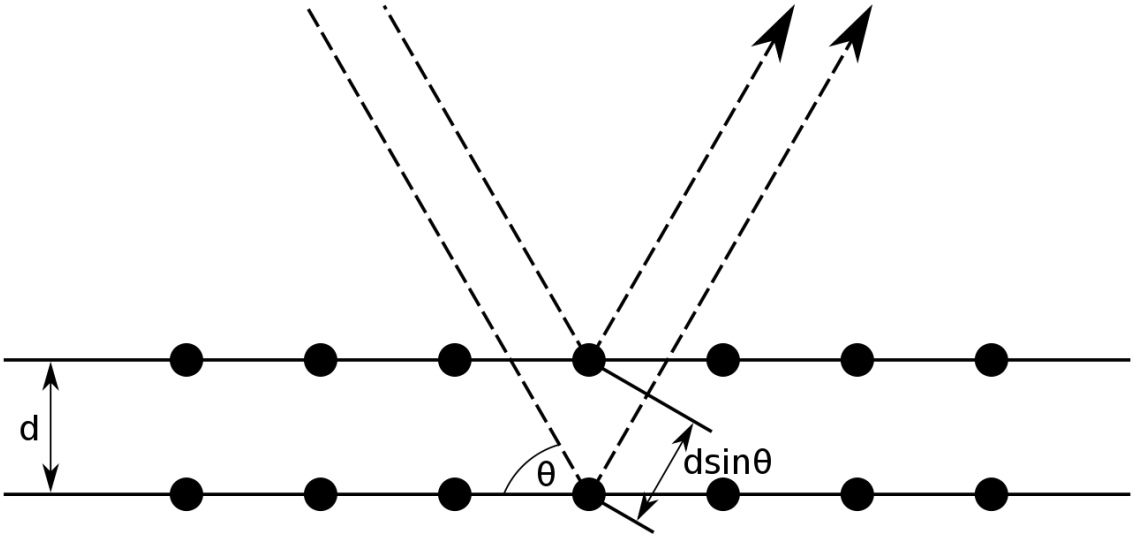


Fig. 3

In the first half of XX century X-rays structure analysis had allowed determination of parameters of crystal lattices of many substances with the help of Eq. (7). Later it was found that spatial patterns of some other small-size objects are also could be stated in this way. For example, double helixes, of which the molecules of DNA consist, have been detected by means of X-rays analysis too.

A problem. Derive the Bragg equation. To determine path difference between two interfering rays at *Fig. 3* use the same approach, which had been applied in consideration of light diffraction in thin films (Lecture 8).

LECTURE 13. A BRIEF INTRODUCTION IN QUANTUM MECHANICS. ELECTRON PARAMAGNETIC RESONANCE. NUCLEAR MAGNETIC RESONANCE.

1. Photoelectric effect and the Compton effect

To the end of XIX century our world seemed to be well enough described by physical theory. Newton's mechanics exactly predicted motion of planets and of any subjects at our Earth. Maxwell's electrodynamics united electric charges' motion with electromagnetic fields, optics based on wave concepts explained phenomena of interference and diffraction and helps to design many useful devices. Among some few significant theoretical questions, which remained incomprehensible, was frequency (or, in other words, wavelength) dependence of radiant emittance of the perfectly black body. To solve this problem Max Planck had to admit that electromagnetic waves are generated as well as absorbed not in continuous flux but by small portions called ***quanta*** (1900). Every **quantum**, according to Einstein's hypothesis, has energy equal to

$$E = h \cdot f \tag{1}$$

where f is the electromagnetic wave frequency, while h is the Planck constant. Light particles, quanta are also named **photons**.

Photoelectric effect was also inexplicable from the classical point of view. It was first observed by Heinrich Hertz in 1887, when he had found that irradiation of the negative electrode with ultraviolet radiation diminished voltage, at which spark discharge began. If cathode was made of an alkali metal, the effect had been observed with visible light. But Hertz did not pay much attention to his discovery, therefore characteristic features of that phenomenon were studied later by other researchers.

An electric scheme used in photoelectric experiments is shown at *Fig. 1*. When the cathode of a vacuum tube is irradiated an electric current flows in the circuit. It has been found that the current vanishes if the wavelength of incident light reaches some threshold value λ_{th} and above this cutoff wavelength no increase of light intensity could generate electric current. Illumination of a positive electrode with the same light does not lead to generation of the photocurrent in the circuit. A typical current-voltage characteristics measured for photoelectric effect is presented at *Fig. 2*.

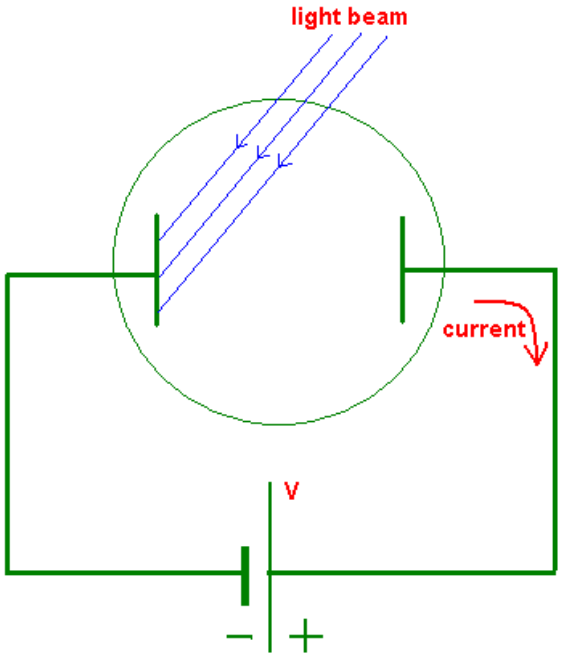


Fig. 1

It is evident form the proposed graph that electric light in the circuit starts from some negative value of anode-cathode voltage, denoted in *Fig. 2* as $-V_0$. At some other value of voltage value of electric current comes to its maximal value I_{sat} (saturation current) and further increase of voltage does not change current. Though it has been found that increase of light intensity (at the same wavelength) increases value of the saturation current.

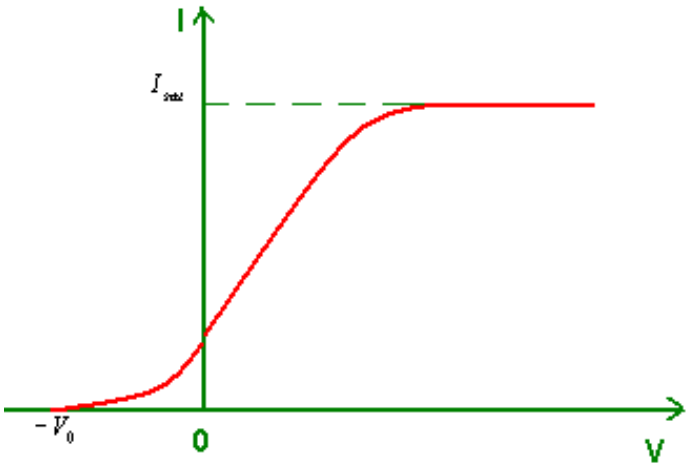


Fig. 2

From the point of view of classical physics electromagnetic wave when being absorbed in a solid could transmit its energy to thermal vibrations of atoms or molecules of the cathode, so that kinetic energy of

electrons exceeds the **work function** A_{wf} of a substance, which the cathode is made of (work function is binding energy of an electron in an atom). But this transmission should not depend on the wavelength of incident radiation. Moreover if classical explanation was correct, the kinetic energy of released electrons would strongly depend on the light intensity. Numerous experiments showed absence of this dependence.

Albert Einstein successfully used Planck's idea of light generated and absorbed in quanta and formula (1) to explain phenomenon of photoelectric effect. He assumed that one ultraviolet quantum acts at only one electron in the anode (this is valid for not very high light intensities), so we can write equation of energy balance

$$hf = A_{wf} + \frac{mv^2}{2} \qquad (2)$$

Photon energy is divided between the work function and kinetic energy of a released electron. It is seen from (2) that in such "one-to-one" interaction photoelectric effect is possible only if energy of the incident photon is more than the work function of a material of anode. Kinetic energy of the electron increases with increase of the photon's frequency, therefore it does not depend on the radiation intensity (number of photons), though number of electrons (saturation current) depends. Evidently that the threshold wavelength from the latter equation is equal to (it corresponds to the frequency, at which all photon's energy is used to overcome electron's binding energy:

$$hf_{th} = \frac{hc}{\lambda_{th}} = A_{wf}$$

Existence of the negative-voltage bias $-V_0$ in *Fig. 2* is also explained by Equation (2): if energy of the quantum of electromagnetic radiation is more than the work function of an anodic substance then excess of electron's kinetic energy can be used for overcoming of negative voltage in the tube.

Another one experiment, which questioned validity of classical description of processes in the microworld had been conducted by Arthur Compton in 1922. He studied scattering of high-energy X-radiation at crystals. In the case of low-energy radiation elastic scattering of X-rays takes place resulting in diffraction pictures of crystals under investigation. Contrary to that Compton observed shift of wavelengths of incident photons depended on the angle of scattering and equal to

$$\lambda' - \lambda = \frac{h}{mc} \cdot (1 - \cos\phi) \qquad (3)$$

where λ and λ' are wavelengths of the incident and of the scattered photons respectively, m is the mass of electron, c is the light speed in vacuum and ϕ is the angle, at which a given quantum is scattered.

Compton derived formula (3) considering also "one-to-one" interaction of the incident photon with an electron of crystal. Part of energy of the photon is transmitted to an electron, so energy of the photon diminishes and therefore increases its wavelength. In fact, Equation (3) was derived using only conservation of momentum and of energy laws.

2. Particle-wave dualism

Thus it had become evident after explanation of the photoelectric effect as well as the Compton effect that light has dual properties – of a wave as well as of a particle.

In 1924 Louis de Broglie assumed that every microparticle like electron or proton also had wave properties and its wavelength had been inversely proportional to its momentum p:

$$\lambda = \frac{h}{p} \qquad (4)$$

This hypothesis was confirmed in 1927 by Clinton Davisson and Lester Germer, who observed experimentally diffraction of electrons at metallic crystals. Using electrons with sufficiently big kinetic energy it is possible to get their wavelengths some orders less than 300-400 *nm* – that is a resolution limit of optical microscopes. This effect is used in electron microscopy. Electron beam is emitted by a heated cathode, which is charged negatively, and moves with acceleration towards a positive anode. In the anode electrode a small orifice is made for flight of electron beam. That beam bombards an object placed after the anode and an image formed by reflected or transmitted electrons is considered at a screen, like we look at the image of an object in an optical microscope, detecting photons reflected from or passes through an object.

In 1970 Merli, Pozzi and Missiroli observed diffraction of electrons at two narrow slits made in opaque plate. The scheme of their experiment is shown at *Fig. 3*: from the source (electron gun) *1* electron beam passes through two thin slits in the plate *2* and diffraction pattern is formed at the screen *3*. The screen could be made from scintillating material to illumine at

points where electros were absorbed. Maximal intensity of diffraction pattern was observed opposite the middle of the distance between slits as it would be in light diffraction but it was rather surprising in the case of propagation of particles. More intriguing results were obtained in experiments carried out by Biberman, Sushkin and Fabrikant in 1948-1949. They used electron beam of so low intensity that electrons were separated from each other by time intervals 104 times longer than time of motion of one electron from the gun to the screen, so in fact only no more than one electron could pass through slits at any moment of time. Nevertheless, diffraction pattern is formed after rather long time by addition of many electrons, it shown that every electron has wave properties but not their set.

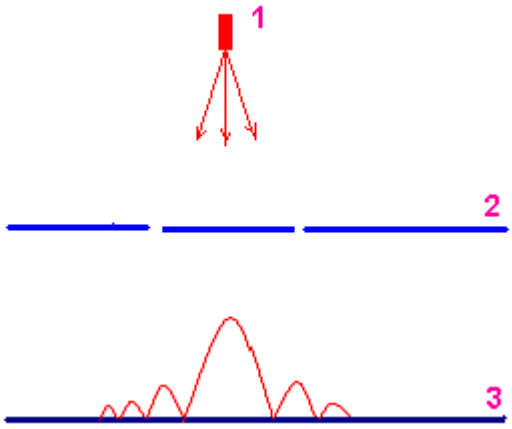


Fig. 3

3. Uncertainty principle

The next very important step in development of quantum mechanics was made by Werner Heisenberg (1927) who formulated the **uncertainty principle**: it is fundamentally impossible to determined exactly values of some pairs of parameters of microparticles. Among those pairs are a coordinate of a particle and the component of its momentum along the same direction, energy of a particle and time of being in the state with that value of energy and some others. Quantitatively, this principle looks like:

$$\Delta x \cdot \Delta p_x \geq \eta, \quad \Delta E \cdot \Delta t \geq \eta, \ldots \qquad (5)$$

where $\eta = h/2\pi$.

Uncertainty principle means that if we enhance precision of measurement of a coordinate, for example, of a particle, unavoidably we

deteriorate result of measurement of its momentum (or velocity, in other words). Manifestation of this principle in classical mechanics ("in ordinary life") is imperceptible because of very small value of Planck constant.

To demonstrate uncertainty principle let us recall that minimal distance distinguishable with the help of a microscope is of the order (an idea of the next two examples is picked up from Leonard Schiff's book "Quantum mechanics")

$$\Delta x \sim \frac{\lambda}{\sin\left(\alpha/2\right)} \tag{6}$$

where λ is a light wavelength, α is aperture angle. (For simplicity we consider experiment being carried out in air with refractive index $n \approx 1$.) Suppose also that, using a microscope, we try to discern some small object with the help of only one photon (light particle). From Equation (4) momentum of that photon is $p = h/\lambda$. But this is its total momentum and photon when passing through the microscope lens could propagate in any direction. The *x*-component of its momentum can be estimated with inaccuracy $\Delta p_x \sim h\sin(\alpha/2)/\lambda$. After interaction of the photon with a particle, which we try to discern, they can exchange with their momenta, therefore inaccuracy of measurement of *x*-component of the particle is of the same order as (4). It means that product

$$\Delta x \cdot \Delta p_x \sim h$$

that is 2π times more than the minimal estimation given by the first inequality of (5).

Let us consider constraints which uncertainty principle imposes on results of our thought experiment on electron diffraction by two parallel slits. Suppose, special detectors are mounted near every slit to register electrons' passage (small rectangles near the slits under the diaphragm *2* at the *Fig.4*). Then uncertainty in determination of position of a detector could not be more than

$$\Delta y = \frac{a}{2} \tag{7}$$

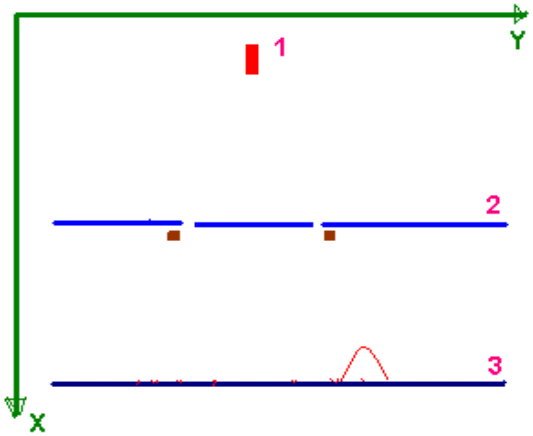


Fig. 4

where *a* is a distance between slits, because a detector is definitely placed behind one slit or behind another. In opposite case it would be impossible to resolve through what slit did electron actually pass. Uncertainty of *y*-component of momentum of electron is limited by the expressions

$$\Delta p_y << \theta \cdot p_x \sim \frac{\lambda}{a} \cdot p_x$$

where θ is the angle between directions to the 0^{th} maximum and 1^{st} minimum from a given slit. The last expression can be rewritten with the help of Eq. (6) provided that x-component of momentum is much more than others, that is $|p| = \sqrt{p_x^2 + p_y^2 + p_z^2} \approx p_x$:

$$\Delta p_y << \frac{h}{a \cdot p_x} \cdot p_x = \frac{h}{a} \qquad (8)$$

The product of expressions (7) and (8) leads to

$$\Delta y \cdot \Delta p_y << \frac{h}{2}$$

The last inequality contradicts uncertainty principle, it means that in case of placing detectors to determine through which slit electron (as well as any other microscopic particle, like photon “realy flights” we loose experimentally diffraction pattern. For example, if the right detector at the Fig.4 establishes that the electron has passed through it, then the image at the scintillating screen 3 is the same like it would be if only right slip is

open while the left one is closed. So if we try to bypass uncertainty principle, nature resists our violation.

4. Schrödinger equation

The main equation of non-relativistic quantum mechanics was formulated by Erwin Schrödinger in 1926. For a particle of mass *m* moving along *x*-axis it stated in a free space (without any external forces)

$$i\frac{\partial\psi}{\partial t} = -\frac{\eta}{2m}\cdot\frac{\partial^2\psi}{\partial x^2} \tag{9}$$

where $i=\sqrt{-1}$, ψ is called the **wave function** of a particle under consideration. Value of a wave function in a given point at a given moment of time is proportional to probability to find the particle in that point at that time. Probability to find some particle somewhere for $-\infty < x < \infty$ is equal to one, if this particle realy exists at a given moment of time therefore:

$$\int_{-\infty}^{+\infty} \left|\psi(x,t)\right|^2 dx = 1 \tag{10}$$

Equation (10) is the so-called **condition of normalization** of a wave function.

The Schrödinger's equation is used to calculate wave functions for particles at different conditions.

It should be emphasized that equation (9) could not be derived from any other equations of classical physics, it has been a breakthrough, a result of generalization of experimental data.

You can easily check that the next solution satisfies the Schrödinger's equation in form (9):

$$\psi = A_0 \exp\left[i(\omega t - kx)\right] \tag{11}$$

if $\omega = \eta k^2 \big/ 2m$. A_0 is an amplitude of the solution, ω is its frequency, while *k* is a wavenumber. If you have never met expression like that on the right-hand side of Equation (11) (it is called the complex exponential function), you can formally treat (11) as any real function $y(x)=\exp(ax)$ when differentiating formula (11) and substituting the result in equation (9). But it

would be useful to know also that as it is stated in the mathematical theory of functions of complex variables

$$\exp[i(\omega t - kx)] = \cos(\omega t - kx) + i\sin(\omega t - kx)$$

therefore we can say that solution (11) represents a wave propagating from left to right along the *x*-axis.

Let us estimate energy of a particle with wave function given by (11). According to the Planck's relation (1)

$$E = h \cdot f \equiv \eta \cdot \omega = \frac{\eta^2 k^2}{2m}$$

We believe like in classical mechanics $k = 2\pi/\lambda$ then substituting Equation (4) into the latter expression for energy we obtain

$$E = \frac{p^2}{2m}$$

that is a well-known expression for kinetic energy in classic mechanics.

In the case of some potential force acting at the particle (a force depending only on distances between objects, like gravitation or electrostatic field)

$$F(x,t) = -\frac{\partial U(x,t)}{\partial x}$$

then Eq. (9) acquires a more general form:

$$i\frac{\partial \psi}{\partial t} = -\frac{\eta}{2m} \cdot \frac{\partial^2 \psi}{\partial x^2} + U(x,t)\psi \qquad (12)$$

Let us apply Equation (12) to the next problem. Consider an electron in a potential well of rectangular form with infinite walls (*Fig. 5*):

$$U(x) = 0 \text{ if } 0 \le x \le L \quad \text{and} \quad U(x) \to \infty \text{ if } x < 0 \text{ or } x > L \qquad (13)$$

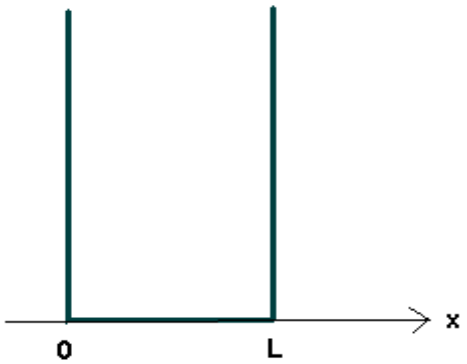


Fig. 5

It can be easily checked that the solution to Equation (12) at the interval $0 \le x \le L$ is

$$\psi = \left(A\cos kx + B\sin kx\right)e^{-i\omega t} \qquad (14)$$

Suppose additionally that $\psi = 0$ at both boundaries of the well (for $x = 0$ and $x = L$), because nothing could penetrate through potential walls of infinite height. Then equating solution (14) to zero at the boundaries of the interval $0 \le x \le L$, we have got

$$A = 0 \quad \text{and} \quad B = 1\,, \ \text{for}\ kL = n\pi\,, \qquad (15)$$

where n is a whole number. So the final form of solution (14) provided (15) is

$$\psi = \sin\left(\frac{\pi n x}{L}\right)\cdot\exp\left(-i\frac{\eta n^2\pi^2 t}{2mL^2}\right) \qquad (16)$$

As we know that the square of absolute value of a wave function determines probability to find particle, it is seen from solution (16) that there are some points within potential well, in which the electron could be never found (where are *sin*-function in (16) is equal to zero) as well as at some other points probability to find it has the maximal value.

Similar problem can be solved for the case of finite depth of a quantum well using a bit more complicated algebra. It was surprisingly found in this way that a particle (e.g. electron) could penetrate through the finite-height walls of a well even if its energy is less than the wells' height, probability of this phenomenon could excide zero. Effect of penetration of microparticles through such potential barriers is called **quantum tunneling**.

5. Electron paramagnetic resonance

A consistent theory of electron paramagnetic resonance (as well as of nuclear magnetic resonance) is essentially based on quantum mechanics and needs knowledge of linear algebra, so we consider only qualitative explanation, which nevertheless can clarify basic principles of these methods.

As you know from the Lecture 6, unpaired electrons of paramagnets induce magnetic moments. In the presence of external magnetic field the magnetic moment of an unpaired electron can be oriented either along the field direction or oppositely to it. Quantitatively, it has been found that every energy level E_0

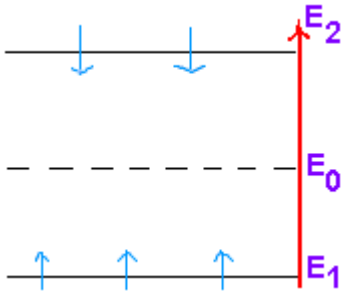


Fig. 6

is splitted in magnetic field into two sublevels E_1 and E_2 (in *Fig. 6*, the long vertical arrow shows direction of external field, short arrows are in the direction of magnetic moments of atoms) so that

$$\Delta E = E_2 - E_1 = 2(E_2 - E_0) = 2(E_0 - E_1) = g\beta H$$

where the constant $\beta = {q_e h}/{4\pi m_e} = 9.27 \cdot 10^{-24}\ {Joule}/{T}$ is called the **Bohr magneton** (recall also the Lecture 6), constant g is called the **Landé splitting factor** (or simply ***g*-factor**).

If paramagnetic material being placed in magnetic field, is exposed to electromagnetic wave, so that energy of some photons is exactly equal to $hf = \Delta E = g\beta H$, then electrons from the upper level pass to the lower one, radiating photons, while electrons from the lower level pass to the upper one, absorbing photons. As electrons in the equilibrium state obey Boltzmann statistics, ratio of electrons' population at sublevels E_2 and E_1 is

$$\frac{n_2}{n_1} = e^{-{\Delta E}/{kT}}$$

so number of transitions in the direction from E_2 to E_1 is much less than in opposite one. Therefore under such conditions paramagnetic substance absorbs energy of incident EM wave. This phenomenon is called **electron paramagnetic resonance** (EMR). EMR is very important tool of investigation of energy states in substances containing free radicals, like intermediate products of chemical reactions, aliovalent metals.

In experimental conditions a spacemen is usually placed between poles of an electromagnet, value of magnetic field of which is gradually changed. When irradiating the spacemen with a microwave (*Fig. 7*), value of magnetic field intensity corresponding to the resonant condition

$$f = \frac{E_2 - E_1}{h} = \frac{g\beta H}{h}$$

is determined experimentally.

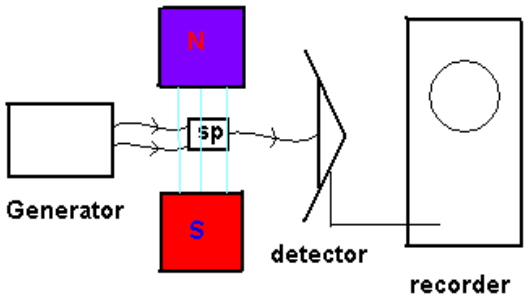


Fig. 7

Resonant frequency (which is proportional to intensity of magnetic field H) can be determined from the position of maximal point at the graph of dependence of absorption coefficient A of EM radiation on H (*Fig. 8, a*). But the dependence of derivative dA/dH on H (*Fig. 8, b*) is used more frequently to increase accuracy of measurements, because the zero point of an experimental curve is much easier to detect than a maximum.

It was found empirically that:

1. Every special kind of free radicals has its own value of the Landé factor, that helps to separate different radicals.

2. Area under the graph of EPR (*Fig. 8, a*) is proportional to the number of free radicals in a specimen.

3. Shape of EPR graph depends on properties of an investigated substance, namely the wider is the EPR curve, the more viscous is this substance and so on.

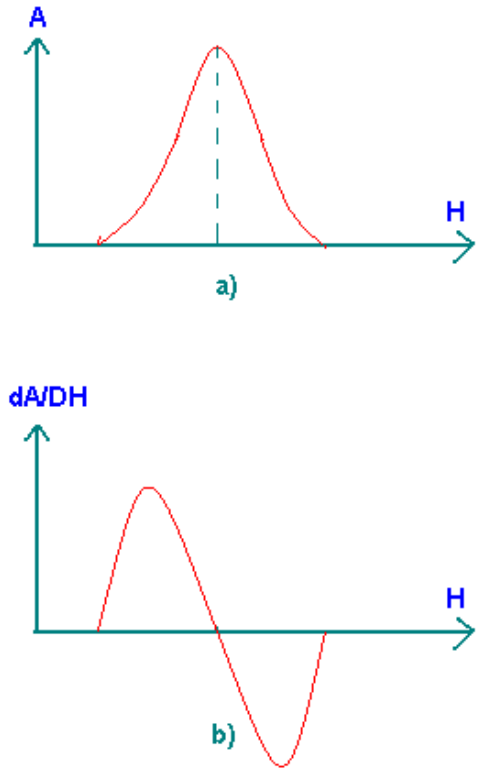


Fig. 8

5. Nuclear magnetic resonance

Atoms with odd number of nucleones have uncoupled nuclear magnetic moment. Total magnetic momentum of a nucleus ${}^{A}_{Z}X$ is equal to sum of all magnetic moments of protons and neutrons:

$$\vec{m} = \sum_{i=1}^{Z} \vec{m}_{ip} + \sum_{j=1}^{A-Z} \vec{m}_{jn}$$

where $\vec{m}_{ip}$ is a vector of magnetic moment of the i^{th} proton, $\vec{m}_{jn}$ is a vector of magnetic moment of the j^{th} neutron.

As it is in the EMR method, energy levels' splitting takes place in presence of external magnetic field:

$$E = E_0 \pm \frac{1}{2} g_n \beta_n H = E_0 \pm \frac{1}{2} g_n \frac{eh}{4\pi m_p} H$$

where g_n is the nucleus g-factor (it has different values for nuclei of different chemical elements) and $\beta_n = 5.05 \cdot 10^{-27}\ Joule/T$ is the **nuclear magneton**. Bohr magneton is far more than the nuclear one because mass of an electron is much less than mass of a proton.

Splitting of energy levels means that measurement of absorption of electromagnetic radiation in a substance in presence of magnetic field can provide information about chemical structure of the substance (**nuclear magnetic resonance** or **NMR** method). It was found that the same atomic nucleus included in bonds with atoms of different chemical elements absorbs EM energy at slightly different frequencies (**chemical shift** of NMR frequency). It makes NMR method a powerful tool of chemical analysis.

LECTURE 14. RADIOACTIVITY. DOSIMETRY.

1. Radioactivity

Natural radioactivity was discovered by Henry Becquerel in 1896, as spontaneous emission by uranium salts of invisible radiation able to ionize air. Ernest Rutherford passed that radiation through the magnetic field and found out that a beam was splitted into three arts. It turned out that trajectory of one beam had changed as it consisted of positively charged particles (Rutherford called them α-particles), another one beam had behaved like a flux of negative charges (β-particles according to his classification) and the third part had propagate in the magnetic field without deflection ($\gamma-$particles). Later it has been found that α-particles were ions of helium, β-particles were electrons and $\gamma-$particles were electromagnetic waves (flow of photons) with frequencies exceeding those of x-rays ($\lambda < 10^{-10}\ m$).

Today we know that an atom of every substance is formed of a positive nucleus and clouds of negative electrons around it. Every nucleus, except for that of hydrogen (which consists from one proton), includes positively charged protons and neutral neutrons. Mass of an electron is $m_e = 9.1095 \cdot 10^{-31}\ kg$, mass of a proton is $m_p = 1836.15 \cdot m_e$, mass of a neutron is $m_n = 1838.68 \cdot m_e$. Number of protons in an atom is denoted with letter *Z* and also called the **atomic number**, because it stated the sequential position of an atom in the periodic table. Sum of all nucleons (protons + neutrons) in an atom is denoted as *A* and also called the **mass number** of an element. Atoms with the same atomic number, but different mass numbers are called **isotopes** of an appropriate chemical element. Mass number is usually marked as the upper index of an element, while atomic number as the lower one. The lower index can be omitted for simplicity, so two notations ${}^{16}_{8}O$ and ${}^{16}O$ mean the same isotope of oxygen.

Atomic nucleus has very small size. In a simple model considering nuclei as spheres estimations state their radii are of the order of

$$r = r_0 \cdot A^{1/3} \qquad (1)$$

where $r_0 \approx (1.2-1.5) \cdot 10^{-15}\, m \equiv (1.2-1.5) \cdot 10^{-2}\ fermi$ (*fermi* is the off-unit measure of length, used in nuclear physics). Even for heavy elements values determined using (1) are much less than radius of even a hydrogen atom, because atoms' radii are mostly defined by the sizes of electron clouds.

Radioactivity is called ability of some nuclei to disintegrate. In the case of **natural radioactivity** they are disintegrated spontaneously. Artificial radioactivity takes place in accelerators under bombardment of atoms with some other particles (nucleons, ions, atoms).

Ability of a nucleus to disintegrate (also called **fissionability**) depends on its structure. The same chemical element can have stable as well as unstable (radioactive) isotopes. For example, there are four stable isotopes of iron: ^{54}Fe, ^{56}Fe, ^{57}Fe, ^{58}Fe and 24 unstable.

2. Radioactive decay law

Radioactivity is a probabilistic phenomenon. Nobody knows, which atom is disintegrating in the next moment of time, but it has been stated that if a sample contains *N* radioactive atoms then in the next short time interval *dt* number of disintegrating nuclei should be proportional to both *N* and *dt*:

$$dN = -\lambda \cdot N \cdot dt \qquad (1)$$

where λ is called the **decay constant** of a given chemical element. Sign "-" points out the decease of radioactive atoms in time.

Equation (1) could be easily integrated yielding the next result:

$$N = C \cdot \exp(-\lambda t)$$

If number of radioactive atoms at the initial moment of time $(t = 0)$ is N_0, then the latter formula comes to

$$N = N_0 \cdot \exp(-\lambda t) \qquad (2)$$

Equation (2) is the (**main**) **law of radioactive decay**. It states time dependence of *nondisintegrated* atoms on time. Often this law is written in a bit different form: the notion of **half-time period** *T* (that is time, in which 50% of radioactive atoms are disintegrated) is used instead of decay constant. Thus

$$\frac{N_0}{2} = N_0 \cdot \exp(-\lambda T)$$

so we can write

$$T = \frac{\ln 2}{\lambda} = \frac{0.693}{\lambda}$$

Finally the law of radioactive decay can be expressed as

$$N = N_0 \cdot \exp\left(-\frac{0.693t}{T}\right) \qquad (3)$$

Half-time period varies from tiny fractions of a second to billions of year for different isotopes of different chemical elements.

Rate of radioactive decay is determined by the **activity**, which is speed of change of number of radioactive nuclei:

$$A = \frac{dN}{dt} \qquad (4)$$

Activity is measured in *becqurels* (*Bq*) in the SI. One *becqurel* is one disintegration per second. There is also an off-system unit *Curie* (*Ci*): 1 *Ci* = 3.7 x 10^{10} *Bq* (this is activity of one gram of ^{226}Ra).

It is evident for (3) and (4) that time dependence of activity is the same as of number of nondisintegrated nuclei.

A problem. Derive the formula of time dependence of activity.

3. Reactions of α- and β- decay

After Ernest Rutherford had separated α-, β- and γ-parts radiation it was found that α-particles and β-particles are products of different processes, which are sometimes accompanied by generation of γ- rays. There are some kinds of decay, such as one-proton decay, two-proton and so on, but we consider here only three basic kinds of reactions.

α-decay is schematically represented as the next reaction:

$$^{A}_{Z}X \rightarrow ^{A-4}_{Z-2}Y + ^{4}_{2}\alpha + \gamma \qquad (5)$$

where *X* is the mother (initial) nucleus, *Y* is the daughter (product) nucleus, γ-particle is denoted with its own Greek letter. Spectrum of α-particles emitting in this kind of nuclear disintegration is discrete and their energy is usually in the range 4-9 *MeV* (*megaelectronvolt*).

There are three different kinds of β- decay: electron one, positron β- decay and electron capture. Their reactions look as follows:

1) A neutron in atomic nucleus in electron β - decay is transformed into a proton, electron and a massless compared with nucleons, uncharged particle called **electron antineutrino** (denoted as $\bar{\nu}$ below):

$$ {}^{1}_{0}n \rightarrow {}^{1}_{1}p + {}^{0}_{-1}\beta + \bar{\nu} \qquad (6) $$

It means that atomic number of an element increases in electron β - decay, while its atomic mass remains the same:

$$ {}^{A}_{Z}X \rightarrow {}^{A}_{Z+1}Y + {}^{0}_{-1}\beta + \bar{\nu} \qquad (6, a) $$

2) In the case of the positron β - decay a mother nucleus emits a positron (the particle-twin of electron but with positive electric charge) and the **electron neutrino**:

$$ {}^{1}_{1}p \rightarrow {}^{1}_{0}n + {}^{0}_{1}\beta + \nu \qquad (7) $$

and

$$ {}^{A}_{Z}X \rightarrow {}^{A}_{Z-1}Y + {}^{0}_{1}\beta + \nu \qquad (7, a) $$

A positron is an example of antiparticle (with respect to an electron). In the case of their collision they annihilate each other and some amount of energy is released in form of γ - radiation.

3) Electron capture is the process of transformation of a proton from a nucleus and an electron from one of inner atomic shells into a neutron combined with emission of an electron neutrino:

$$ {}^{1}_{1}p + {}^{0}_{1}\beta \rightarrow {}^{1}_{0}n + \nu \qquad (8) $$

and

$$ {}^{A}_{Z}X + {}^{0}_{1}\beta \rightarrow {}^{A}_{Z-1}Y + \nu \qquad (8, a) $$

It was stated theoretically that it is presence of neutrino and antineutrino in products of reactions of β - decay that forms continuous spectrum of these kinds of reactions.

4. Interaction of products of radioactive reactions with matter

Initial energy of products of radioactive decay (α-particles**,** β-particles, γ- rays and so on) exceeds ionization potential (binding energy) of any electron in atoms. Therefore their action at a substance is based on ionization of atoms and molecules. Because of ionization products of decay loose their energy, therefore velocity of α-particles or β-particles diminishes, but wavelength of γ- photons increases in propagation through a matter.

To characterize interaction of products of radioactive reactions with matter the next thre values are used.

1) **Linear stopping power (*S*)** is energy lost by an ionizing particle at the unit track length:

$$S = \frac{dE}{dx} \tag{9}$$

Linear stopping power depends on a substance as well as on particle's charge and energy. Usually it is measured in off-system units $keV/\mu m$.

2) **Linear specific ionization (*i*)** is the number of ions pairs, which are generated at the unit track length:

$$i = \frac{dn}{dx} \tag{10}$$

It is measured in number of pairs per micron. If E_i is an average energy of ionization of atoms of a substance, then

$$i = \frac{S}{E_i}$$

3) **Mean linear range (*R*)** is an average path, which ionizing particle passes in matter until loss of kinetic energy.

α-particles usually propagate through matter rectilinearly, first its specific ionization slowly increases, but as its mean linear range is achieved, specific ionization dramatically falls down to zero level. Electrons knocked out from atoms in this way usually cause secondary ionization as newborn β-particles. As kinetic energy of the α-particle reaches the order of temperature value, it catches two electrons and becomes a helium atom.

Average values of mean linear range of α-particles in living tissues are of the order of dozen microns.

β-particles when moving in substances collide with bound electrons of atoms and due to their interactions often change direction of propagation. Linear specific ionization of β-particles also at first grows as they propagate and loose their kinetic energy (they "have more time" for interactions). Depending of initial kinetic energy mean linear range of β-particles varies from fractions of microns to some millimeters.

γ-rays when propagating through matter obey Bougouer law of absorption as other electromagnetic waves. If their energy is sufficient to cause photoelectric effect or Compton scattering, they emit some amount of electrons, that result in secondary ionization with β-particles.

5. Natural radioactivity

Natural radioactivity is a set of processes of disintegration of radioactive isotopes of different chemical elements, which takes place in nature. It was found that four isotopes being contained in the Earth's crust in abundance $^{235}U, ^{238}U, ^{232}Th, ^{237}Np$ are the starting points of appropriate 4 series of radioactive transformations, which are ended with the stable isotopes of lead (three first mentioned series) or bismuth (the last series).

Decomposition products of the series, starting from ^{238}U) are the main sources of internal irradiation of a human body. About half of natural radiation falls on heavy inert gas ^{222}Rn (40% of natural radiation). It is generated in the crust and further leaks upwards. ^{210}Pb and ^{210}Po are gathered in fish and seafoods and are consumed later.

There is also some amount of natural isotope ^{40}K in the Earth's crust. It goes in a human organism with fresh fruits, vegetables and mushrooms (the latter three elements bring about 15% of natural radiation).

An important source of natural radioactivity is cosmic radiation. It includes ionizing particles and electromagnetic waves generated by our Sun and some distant stars (12% of natural radiation).

6. Radiation doses

Notion of radiation dose is used for quantitative estimation of amount of ionizing particles or γ-rays absorbed with different objects.

Absorbed dose (*D*) is energy of ionizing radiation transmitted to unit mass of an object:

$$D = \frac{dE}{dm} \tag{11}$$

The SI-unit of absorbed dose is *gray* (abbreviation is *Gy*). Practical off-system unit is *rad*:

$$1Gy = 1\frac{J}{kg} = 100rad$$

Absorbed dose is mostly used for inanimate substances, because the same energy brought to living organisms by different products of radioactive reactions lead to different aftermath.

Therefore a parameter called **quality factor (*K*)** is introduced as generalization of observation of biological action of radioactive products. It is assumed now that if quality factor of electromagnetic radiation and β-particles to set value 1, then for α-particles it is equal to 20.

Quality factor is a dimensionless value, which is used for estimation of **equivalent dose** of radiation in the next way:

$$H = K \cdot D \tag{12}$$

In fact, equivalent dose could be also measured in grays, though in the SI a separate unit is introduced *sievert* (*Sv*). An off-system unit is *rem* (*Röntgen-equivalent-man*):

$$1Sv = 100rem$$

It was found experimentally that different biological tissues suffer differently from the same amount of even the same radiation. Therefore **weighting factors** were introduced for different human organs (see the table below).

tissue	**weighting factor, k_T**	**tissue**	**weighting factor, k_T**
red bone marrow	0.12	liver	0.05
gonad	0.2	gullet	0.05
narrow intestine	0.12	thyroid gland	0.05
lung	0.12	skin	0.01
bosom	0.12		

Weighting factors are used for calculation of **effective radiation dose**, that is sum of equivalent doses obtained by different organs multiplied by their weighting factors:

$$H_{eff} = \sum_i K_i \cdot H_i \qquad (13)$$

Thus, to calculate effective dose obtained with some patient you should estimate equivalent dose fallen at every organ, multiple this value by the corresponding weighting factor and summarize all those products.

SI-units for the effective doses are the same as for equivalent ones until now (fortunately).

Absorbed, equivalent and effective doses of ionizing radiation characterize values of transmitted energy. Ionizing ability of radiation is determined with the help of **exposure dose (*X*)** obtained by an object. It is determined as total electric charge of all positive ions generated in unit mass of air at normal conditions.

SI-unit of exposure dose is *Coulomb* per *kilogram*. Off-system unit is called *Röntgen* (*Roentgen,* abbreviation is *R*):

$$1R = 2.58 \cdot 10^{-4} \ \frac{C}{kg}$$

If exposure does is equal to 1 *R* then $2.08 \cdot 10^9$ positive ions are generated in 1 cm^3 of the dry air.

It was found that absorbed dose measured in *rad* is approximately equal to exposure dose in *Röntgen* in many substances, such as water and soft tissues of human organisms.

Influence of ionizing radiation on living tissues depends on its intensity and duration. It is known that equivalent dose 5-10 *rem* leads to some mutations, 50-100 *rem* disturbes immunity, 400-500 *rem* guarantees damage of bone marrow and death in 50% cases and so on.

7. Dose limits

Dose limit is annual effective dose of ionizing radiation, which should not be exceeded for safety of people. In the independent states arisen at the place of the former Soviet Union two scales of dose limits are fixed. One scale is for personnel working with dangerous substances (group A), another one scale is for personnel being influenced by radioactive substances in their work (group B). All dose limits for group B are 4 times

less than for the group A. For inhabitants of territories where there is risk of radioactive contamination dose limits are 10-20 times less than for personnel of group A.

Thus, it is stated that the group A personnel should not get more than 50 *mSv*/year or less than 20 *mSv*/year in any 5 successive years. The same values for population are equal to 5 *mSv*/year or 1 *mSv*/year in any 5 successive years.

It is believed that natural radioactive background less than 2 *mSv*/year (that corresponds to exposure dose of the order of 10-20 $\mu R / hour$) is acceptable.

It is clear that to diminish influence of ionizing radiation on people it is necessary to decrease exposure time and/or to place radioactive materials as far from people as it is possible. In case of need of work under action of penetrating radiation there can be used protective screens.

Even a thin paper sheet defends from α-particles.

Thin layer of aluminium or glass with thickness of some centimeters defend from β-particles.

A lead plate (some centimeters) or a layer of concrete of 1.5 –2 meters thickness provides a good defence from γ- rays.

Neutrons are effectively detained with the help of substances containing large amount of hydrogen, like water. (Interaction of a neutron with a hydrogen atom ${}_1^1H$ results in emerging of stable deuterium ${}_1^2H$ or radioactive, but long-lived tritium ${}_1^3H$).

8. Detectors of ionizing radiation

Devices used to register radioactive particles are called detectors. They are divided into radioactive radiation **counters** and **track detectors**.

A counter generates electric pulse when a charged particle is got with it. Counters are divided into counting ionization chambers, proportional counters, Geiger-Müller (GM) counters semiconductor counters and others.

Ionization chambers are used since Ernest Rutherford's time and remain popular until now. A typical chamber consists of a condenser filled with a gas. Voltage of the order of 100 – 1000 *V* is applied to its plates. α- or β-particle, or a proton caught with volume of this condenser generates flow of electrons to a positive electrode and generates a pulse of electric current in the chamber's circuit. This current (or voltage at a resistance in the circuit) is very small usually, so an amplifier is necessary to register the signal from one particle. Electromagnetic waves cannot be detected in such

chamber, but propagation of high-energy γ-radiation in gas is accompanied with generation of secondary particles, which are registered.

Bubble chambers, Wilson cloud chambers, thick layers of photoemulsion and some other devices are referred to track detectors. The oldest detector of this kind is Wilson chamber. Its operation is based on condensation of supersaturated water or spirit vapor at ions generated along the trace of a charged particle. The interior of a chamber is illuminated and photographed from several points, that gives stereoscopic images of tracks. A kind and properties of a particle are determined when studying its path length, thickness of the track and direction of deflection in magnetic field. Momentum of a particle and sign of its electric charge are determined in this way. Principle of operation of the cloud chamber is illustrated at *Fig. 1* (Photograph from the Wikipedia was made by Carl Anderson).

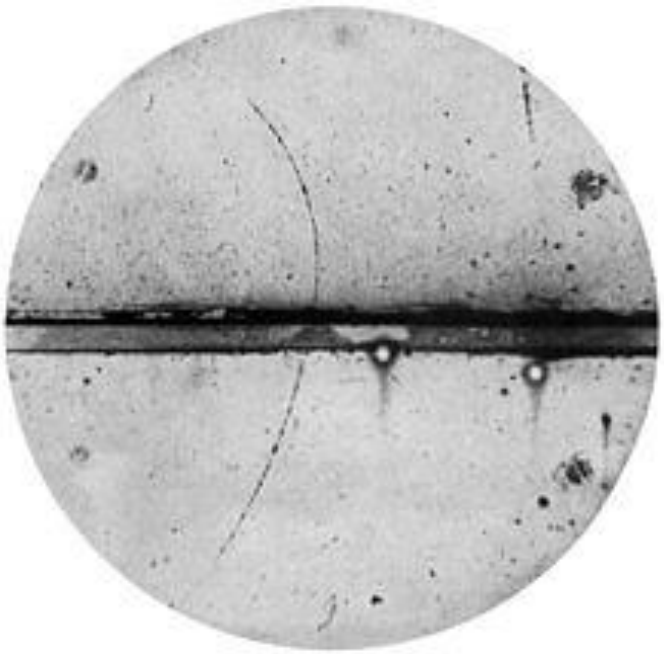


Fig. 1

Positron was theoretically predicted by Paul Dirac. Later Carl Anderson, the author of a photo at *Fig. 1*, discovered it in 1932, as he had considered curved tracks of particles in a Wilson chamber.

Combination of particle detectors with computers allows to process automatically bulk experimental data in our days.

ЛИТЕРАТУРА

1. Сивухин Д.В. Общая физика, тт. 2 – 5. – М., Наука, 1975 – 1988.
2. Баранов А.П., Клименок М.Ф. Медицинская и биологическая физика. – Витебск, ВГМУ, 2003.
3. Федорова В.Н., Фаустов Е.В. Медицинская и биологическая физика. – М., ГЭОТАР-Медиа, 2010.
4. Калитеевский Н.И. Волновая оптика. – М., Высшая школа, 1978.
5. Шифф Л. Квантовая механика. – М., Изд-во иностранной литературы, 1959.
6. Brown B.S. Biological membranes. – London, The British Biochemical Society Publishing, 1996.

Учебное издание

Козлов Александр Игнатьевич

БИОЛОГИЧЕСКАЯ ФИЗИКА

BIOLOGICAL PHYSICS

For the 1st-year foreign students of a pharmaceutical faculty

Пособие

Редактор А.И. Козлов

Подписано в печать _____ Формат бумаги 64×84 1/16
Бумага типографская №2. Гарнитура ТАЙМС. Усл. печ. листов_____
Уч.-изд. л. _____ Тираж _____ экз. Заказ №_____

Издатель и полиграфическое исполнение:
УО «Витебский государственный медицинский университет»
Лицензия ЛП № 02330/453 от 30.12.2013
пр. Фрунзе 27, 210009, г. Витебск